

# ***Essays on Applied Microeconomics***

## D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor rerum politicarum

(Doktor der Wirtschaftswissenschaft)

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

von

M.Sc. Marica Valente

Präsidentin der Humboldt-Universität zu Berlin:

Prof. Dr.-Ing. habil. Dr. Sabine Kunst

Dekan der Wirtschaftswissenschaftlichen Fakultät:

Prof. Dr. Daniel Klapper

Gutachter:

1. Prof. Ph.D. Bernd Fitzenberger
2. Prof. Ph.D. Jeffrey M. Wooldridge

Tag des Kolloquiums: 19.November 2020

*To Giovanna*

## Acknowledgements

I gratefully acknowledge that I was a Ph.D. student of the DIW Berlin, chosen by Helmut Lütkepohl, a great mentor and friend, and further supported by Georg Weizsäcker, who improved our Graduate Center significantly. I thank my first supervisor Bernd Fitzenberger, who always pushed me to be and do better. I thank my second supervisor Jeff Wooldridge, who believed in me since the beginning. I thank Lorenzo Trapani, with whom I share the unique passion for etymons, drinks and metrics. I thank my department and Claudia Kemfert for supporting my endless travels. I thank Stefan Seifert, for our paper and the beautiful years together. I thank my friends, you know why. I thank my office mate, Fabian Stöckl, the king of the Alps. I thank my flatmates, Martin Harding and Max Schäfer, for being there, always. I thank love, for being there always too. I thank my biggest passions, football and cycling, sources of energy and inspiration.

I dedicate this thesis to my mum, Giovanna, for supporting me and to my dad, Angelo, for supporting her in supporting me.

## Summary

In economics, researchers use a wide variety of strategies for attempting to draw causal inference from observational data. New developments in the causal inference literature focus on the combination of predictive methods and causal questions. These methods allow researchers to answer new research questions as well as provide new opportunities to address older research question in the literature. This dissertation entails empirical work in the fields of (i) environmental economics: I evaluate waste pricing policies using synthetic controls and machine learning methods; (ii) labor and migration economics: I identify and quantify unreported farm labor induced by a sudden migrant inflow; (iii) conflict economics: I evaluate the economic costs of an hybrid war, namely, the Donbass war in Ukraine. The contribution of this dissertation is threefold. First, I combine novel data sources and provide unique datasets. Second, I apply and tailor modern evaluation methods to the estimation of policy-relevant causal parameters in various fields of economics. Third, I compare recent versus traditional econometric approaches previously employed by the literature. My dissertation shows that modern econometric techniques hold great promise for improving the accuracy and credibility of causal inference and policy evaluation.

## Deutsche Zusammenfassung

In der ökonomischen Forschung wird eine Vielzahl von Strategien verwendet, um zu versuchen kausale Schlussfolgerungen aus Beobachtungsdaten zu ziehen. Neue Strömungen in der Literatur zu kausaler Inferenz konzentrieren sich auf die Kombination von Methoden zur Vorhersage und kausalen Fragestellungen. Diese neuen Methoden ermöglichen es neue Forschungsfragen zu beantworten und bieten die Möglichkeit bestehende Forschungsfragen in der Literatur neu zu adressieren. Diese Dissertation umfasst empirische Arbeiten in den Bereichen (i) Umweltökonomie: Ich evaluiere die Preispolitik für Abfälle mithilfe der “synthetic control” Methode und Methoden des maschinellen Lernens; (ii) Arbeits- und Migrationsökonomie: Ich identifiziere und quantifiziere nicht gemeldete landwirtschaftliche Arbeitsleistung, die durch einen plötzlichen Migrationszustrom verursacht wird; (iii) Konfliktökonomie: Ich analysiere die wirtschaftlichen Kosten eines hybriden Krieges, des Donbass-Krieges in der Ukraine. Der Beitrag dieser Dissertation zur bestehenden Literatur ist dreifach. Erstens kombiniere ich neuartige Datenquellen und stelle neue Datensätze bereit. Zweitens verwende ich moderne Evaluierungsmethoden und passe sie an, um politisch relevante kausale Parameter in verschiedenen Bereichen der ökonomischen Forschung abzuschätzen. Drittens vergleiche ich neuere mit traditionellen ökonometrischen Ansätzen, die zuvor in der Literatur verwendet wurden. Meine Dissertation zeigt, dass moderne ökonometrische Techniken vielversprechend sind, um die Genauigkeit und Glaubwürdigkeit von kausalen Schlussfolgerungen und die Evaluierung von Politikmassnahmen zu verbessern.

# Table of contents

## General Introduction

Page 7

## Chapter 1.

The effects of pricing waste generation: A synthetic control approach

DOI: <https://doi.org/10.1016/j.jeem.2019.06.004>

Page 9

## Chapter 2.

Heterogeneous effects of waste pricing policies

Page 37

## Chapter 3.

The Arab Spring migrant wave and illegal labor on vineyards:  
Counting the uncountable

Page 91

## Chapter 4.

The economic costs of hybrid wars: The case of Ukraine

DOI: <https://doi.org/10.1080/10242694.2020.1791616>

Page 128

## General Introduction

This thesis addresses various economic problems in the field of environmental economics, conflict economics, labor economics, and migration. It applies state-of-the-art econometrics and machine learning algorithms, and it shows how these techniques could improve over traditional methods of causal inference.

*Chapter 1.*—“The effects of price wasting generation: a synthetic control approach” is coauthored with Matheus Bueno, Ph.D. student of the Toulouse School of Economics, and published in the *Journal of Environmental Economics and Management*. We evaluate the causal effects of unit pricing policies on municipal solid waste. Using a unique panel dataset of Italian municipalities and the case study of Trento, we show that the policy was effective, decreasing total waste and promoting recycling. Importantly, we show how using the synthetic control method could improve over traditional difference-in-differences approaches when unobservables (such as environmental awareness) have time-varying effects on waste generation.

*Chapter 2.*—“Heterogeneous effects of waste pricing policies” builds on the first paper and studies the heterogeneous effects of unit prices on waste demands and social welfare. The topic is highly non-trivial, especially because of the large dimensionality of the characteristics which are bound to affect the effects of waste policies, the consequent heterogeneity, and the data collection and preparation. I employ machine learning techniques on a unique panel of Italian municipalities with large variation in prices and observables. I show that prices as well as socio-economic attributes matter, and trigger different policy responses. Despite the large effect heterogeneity, I estimate overall welfare benefits after three years of adoption, when waste prices cause significant waste avoidance. This implies that even low prices can change waste behaviors and improve welfare.

*Chapter 3.*—“The Arab Spring migrant wave and illegal labor on vineyards: Counting the uncountable” is written with Stefan Seifert, Post-doc of the University of Bonn, and published in the *DIW Discussion Papers*. We provide a framework to detect illegal labor from official statistics. We exploit an exoge-

nous shock on vineyards, in particular, the 2011 Arab Spring migration wave. We show that illegal labor can be identified and quantified using observed farm data and a dynamic panel model. This article is one of the few academic study of labor practices that are pervasive not just in winemaking, but in all of agriculture.

*Chapter 4.*—“The economic costs of hybrid wars: The case of Ukraine” is co-authored with Julia Bluszcz, Master student at the Humboldt University of Berlin, and published in *Defense and Peace Economics*. We analyze the economic costs of the Ukrainian conflict in the Donbass region, using again the synthetic control method. Between 2013 and 2017, we estimate an average GDP loss of 15% due to the war. However, starting from 2016, the Ukrainian economy showed signs of recovery, a growth that slowly continues driven by external and internal factors, such as world raw material prices, the implementation of macroeconomic and structural reforms, and the improved political situation.



# The effects of pricing waste generation: A synthetic control approach

with Matheus Bueno\*

*Published in the Journal of Environmental Economics and Management here.*

**Abstract** To internalize pollution externalities into household waste generation, Unit Pricing Systems (UPS) have been adopted worldwide. This paper evaluates the causal effects of a UPS on the disposal of municipal solid waste in Trento, Italy. Using a unique panel dataset of monthly waste generation in Italian municipalities, we employ the synthetic control method, which allows us to account for possible time-varying effects of unobservables. Our results show that the policy was effective, with a significant decrease of the priced waste stream, unsorted waste, by 37.5%. This effect seems to be largely driven by behavioral changes towards waste avoidance ( $-8.6\%$ ) and possibly by a smaller increase in recycling ( $+6.1\%$ ). By comparing these results to those obtained by a difference-in-differences approach, we show that failing to account for time-varying effects of unobservables may lead to a mismeasurement of policy effects.

**Keywords:** Waste generation, Unit pricing, Synthetic controls, Policy endogeneity, Selection on unobservables

**JEL Codes:** D01, C21, Q53

---

\*Toulouse School of Economics, Toulouse, France. Email: matheus.bueno@tse-fr.eu.

# Introduction

The process involving generation, collection and disposal of Municipal Solid Waste (MSW) characterizes a classic example of a negative externality, as there are environmental and treatment costs associated with MSW that add to the individual’s private costs. Standard flat fees on MSW collection, however, do not usually suffice for internalizing the cost difference into individuals’ waste generation behavior. Therefore, pricing waste per collected unit, a form of Pigouvian fee (Pigou, 1932), has been increasingly used as a policy instrument, known as Unit Pricing Systems (UPSs) or pay-as-you-throw programs. Existing empirical evaluations of such policies generally find them to be effective, when controlling for selection bias due to unobserved determinants of their adoption with time-invariant effects on waste generation.<sup>1</sup>

Nevertheless, despite their growing popularity, there is an empirical gap on UPSs’ effectiveness when the effects of unobservables on waste generation vary over time, and on the behavioral mechanisms behind UPSs’ effects. Using a unique data on monthly waste generation for nineteen municipalities in Italy between 2008 and 2016, this paper attempts to fill this gap by evaluating the causal effects of a UPS implemented on the disposal of unsorted MSW in the municipality of Trento in 2013.

The first contribution of this paper is to address the endogeneity associated with UPS adoption due to selection on unobservables with time-varying effects. Waste generation in municipalities adopting this policy may have unobserved determinants that spawn diverging waste generation trends with respect to municipalities not adopting it. For example, while environmental consciousness may be fixed over time, the growing concern over the negative environmental and health impacts of waste may lead individuals to increase recycling or curb waste generation (see, for instance, EU’s awareness-raising reports by O’Leary et al. 2017 and European Commission 2010). The impacts of unobservables on waste generation may cause non-parallel trends in outcomes, invalidating the identification assumption of standard policy evaluations based on differencing out fixed individual characteristics with time-invariant effects, such as with fixed effects and difference-in-differences (DID) estimation (Gobillon and Magnac, 2016).

To account for time-varying effects of unobservables, we implement the

---

<sup>1</sup>See, e.g., the survey in Huang et al. (2011), and the literature review in Section 1.

Synthetic Control Method (SCM) (Abadie and Gardeazabal, 2003; Abadie et al., 2010). This method reconstructs the counterfactual outcomes using a combination of untreated municipalities with similar outcome trajectories that did not select into treatment, hence replicating the unobserved heterogeneity and allowing it to have time-varying effects. Several empirical applications in other areas of study have been implementing the SCM<sup>2</sup>, but, in the authors’ best knowledge, this is the first paper doing so to evaluate waste pricing policies.

Secondly, this paper adds to the literature by studying the behavioral mechanisms behind the partial effects on the priced waste stream. When the relative price of unsorted waste increases, individuals may shift away from generating unsorted waste either by recycling more or by avoiding waste generation of any sort, e.g., by using less packaging.<sup>3</sup> Thereby, this paper looks at which behavioral response prevails by evaluating policy causal effects separately for total, recycling and unsorted waste.

Our results indicate policy effects of sizable magnitude, and changes in waste generation behavior. Unsorted waste generation fell immediately after UPS implementation and levelled off after around two years, with an average decrease of 37.5% in the post-treatment period. This partial effect seems to have been largely driven by waste avoidance and a relatively smaller increase in recycling, with total waste decreasing by 8.6% and recycling waste increasing by 6.1%. Placebo tests show that the unsorted waste reduction is strongly significant, while weaker and no significance is found for the causal effects on total and recycling waste, respectively. Finally, we show that the use of the SCM rather than a DID approach is justified in this application because the parallel trend assumption needed for the validity of the latter is violated. The successful reconstruction of the unobserved heterogeneity by the SCM, on the other hand, allowed to control for its time-varying effects.

The remainder of the paper is organized as follows: Section 1 provides a brief review of the related literature on UPS, and discusses potential methodological shortcomings. Section 2 outlines the SCM used to estimate the policy causal effects under study. Background and data are described

---

<sup>2</sup>See, e.g., Reimer and Haynie (2018), Corral and Schling (2017), Almer and Winkler (2017), Acemoglu et al. (2016), Abadie et al. (2015), Kirkpatrick and Bennear (2014), Billmeier and Nannicini (2013), Cavallo et al. (2013), and Pinotti (2015).

<sup>3</sup>Note that a potential policy response is also to dump waste illegally. We discuss this possibility in Section 4.

in Section 3. Section 4 provides the results, and Section 5 concludes and discusses implications of this study.

## 1 Literature

Starting in the 1970s, UPSs have been broadly implemented both in Asia and the United States, e.g., in municipalities of Japan, Korea, California, and Michigan (Sakai et al., 2008; Miranda and Bauer, 1996). By the mid-2000s, UPSs have been used in 30 of the 100 largest municipalities in the country as, e.g., Seattle and San José (Skumatz, 2008). Afterwards, UPS was adopted in many European countries, particularly Switzerland, the Netherlands, the northeastern area of Germany, Denmark and Italy (for a review, see, e.g., Reichenbach, 2008).

Since the 1990s, following their implementation, UPSs' effects have been extensively studied in the literature. Evaluations in the United States respond for a large chunk of the early and current literature (Fullerton and Kinnaman, 1996, 2000; Hong et al., 1993; Huang et al., 2011; Miranda and Bauer, 1996; Podolsky and Nestor, 1998; Reschovsky and Stone, 1994; Van Houtven and Morris, 1999; Wright et al., 2018). Studies for other regions have followed through for, e.g., Korea (Hong and Adams, 1999; Kim et al., 2008), Japan (Usui, 2009; Usui and Takeuchi, 2014; Yamakawa et al., 2002), and the Netherlands (Allers and Hoeben, 2010; Dijkgraaf and Gradus, 2009, 2004; Linderhof et al., 2001). However, to the authors' knowledge this is one of the first UPS policies to be evaluated in Italy and Europe, except from the Dutch case, another study on Italy (Buccioli et al., 2015) and Switzerland (Carattini et al., 2018). The scarce empirical evidence on the effects of European UPSs lingers despite its increasing use in the region, and the EU mandate targeting both waste avoidance and diversion (EEA, 2009, 2013; EU, 2008).

Moreover, results on UPSs' effectiveness differ across studies and geographic areas, suggesting a general lack of external validity. A potential reason for this is that policy effectiveness is likely to vary across different social environments (see, e.g., Kipperberg, 2007; Reschovsky and Stone, 1994, for a thorough analysis). Based on economic theory, one would expect that households respond to economic incentives in a similar way, i.e., by decreasing (increasing) priced (unpriced) waste generation. Empirically, instead, some studies estimate small, often insignificant substitution between

priced and unpriced waste generation, as well as unclear prevention efforts, i.e., waste reductions possibly due to reuse and adjustments in purchasing habits (see, e.g., Allers and Hoeben, 2010; Fullerton and Kinnaman, 2000, 1996; Jenkins et al., 2003).

A challenge that the literature faces is to address the endogeneity of UPS adoption (see, e.g., a discussion in Dijkgraaf and Gradus, 2009). As mentioned earlier, a UPS may be implemented in municipalities where environmental awareness has dynamic effects on waste generation through, e.g., learning effects from additional information provision, and recycling and waste reduction promotion. On the contrary, UPS may also be adopted where low learning effects and environmental awareness result in high waste levels requiring local governments to consider policy alternatives to standard flat fees. Hence, policy adoption is likely not orthogonal to municipalities' observed and unobserved waste generation determinants. To the best of our knowledge, the existing literature only considers selection bias due to observables and/or fixed unobservables with time-invariant effects.

To account for unobserved heterogeneity, various approaches have been used. A part of the literature that controls for observed determinants within a regression framework to estimate price elasticities of waste demand finds them to be mostly inelastic (see Huang et al., 2011, for a discussion of these studies). Among this literature, a number of studies explicitly take into account selection biases with instrumental variable approaches (see, e.g., Fullerton and Kinnaman, 2000; Usui, 2008; Huang et al., 2011). Most of these evaluations are based on cross-sectional data and on a large number of UPS municipalities with different price levels which are instrumented in 2SLS regression (for a review, see Bel and Gradus, 2016). Findings from this literature generally vary, depending on the method, data and municipalities under study.

Other investigations account for unobserved heterogeneity with time-invariant effects in DID and fixed effects estimations, concluding that UPSs are mostly effective in reducing unsorted waste and increasing recycling (Dijkgraaf and Gradus, 2004, Van Houtven and Morris, 1999, Linderhof et al., 2001, Allers and Hoeben, 2010, and Usui and Takeuchi, 2014,<sup>4</sup> Carattini et al., 2018). While controlling for time-invariant unobservables can be an improvement over previous analyses, DID (fixed effects) methods do not

---

<sup>4</sup>Note that Allers and Hoeben (2010) and Usui and Takeuchi (2014) also tested the use of instruments to correct for the endogeneity of the price.

allow the impacts of unobservables on waste generation to vary over time. In fact, although DID allows for different outcome levels between units, it assumes that outcome differences between UPS and non-UPS municipalities do not change over time.

However, in some applications, it can be the case that unobserved determinants of waste generation do not only lead to level differences in waste generation, but have also time-varying effects (as discussed in, e.g., Usui and Takeuchi, 2014). Miranda and Bauer (1996) report that UPS municipalities often engage in citizen education efforts to strengthen environmental awareness, informing households on the excess waste problem with informative campaigns. This seems also the case for the municipality of Trento, where, over the years, citizens and retailers were involved in education and information programs promoting recycling and waste avoidance (Comune Trento, 2012).<sup>5</sup>

Therefore, waste generation in UPS municipalities may vary over time depending not only on the above-mentioned learning effects but also on the target and the intensity of such awareness-raising programs. Indeed, pro-environmental behavior seems to be affected by the ability of local governments in motivating civic mindedness, i.e., the sense of voluntary public good provision through a shared sense of obligation (Alesina et al., 2017; Besley and Reynal-Querol, 2011). Further, also the availability of time and space may impact waste generation of UPS municipalities differently over time, altering the opportunity costs of waste reduction and recycling.<sup>6</sup>

Hence, if determinants of selection bias have time-varying effects on waste generation, failing to account for them would return biased estimates of UPS' true impacts. In this direction, the SCM allows for time-varying effects of the unobserved determinants by recreating them in the treated municipality with a combination of untreated municipalities. This generalizes the DID estimation, as shown by Gobillon and Magnac (2016).

Finally, the literature provides scant empirical evidence on the behavioral mechanisms behind the reduced-form estimated effects of UPS implementations. Theoretical studies point towards both monetary and non-monetary incentives as important determinants of household waste genera-

---

<sup>5</sup>As a result, in some years, the National Institute of Statistics ranked Trento as one of the most environmental-friendly municipalities for, e.g., recycling rates, green spaces, air pollution, energy consumption, and public transport (ISTAT, 2011; La Repubblica, 2006).

<sup>6</sup>For example, time (space) opportunity costs of recycling refer to the time spent (space needed) to sort and transport (store) recycling materials.

tion behavior: While recycling waste or avoiding its generation are associated with opportunity costs in terms of time and effort spent, its benefits involve not only saving eventual fees on waste disposal but also the psychological reward, aka “warm-glow”, associated with environmentally friendly attitudes (see, e.g., Bénabou and Tirole, 2003; Brekke et al., 2004; Jenkins et al., 2003; Kahn, 2007; Morris and Holthausen Jr., 1994; Thøgersen, 2006). Thus, the direction and relative strength of UPSs’ effects likely varies with levels of intrinsic motivation, recycling and avoidance habits as well as socio-economic characteristics such as education, income, and age. On the one hand, pricing waste per unit could lead to increased recycling and waste avoidance. On the other hand, waste generation may stagnate or even increase due to the crowd out of intrinsic motivations (Abbott et al., 2013; Bruvold and Nyborg, 2003). This paper analyzes these behavioral responses empirically by estimating causal effects of UPS on all waste streams: unsorted, recycling, and total.

## 2 Methodology

Instead of differencing out fixed unobserved heterogeneity, the synthetic control method, introduced by Abadie and Gardeazabal (2003), attempts to replicate the unobserved heterogeneity by matching the observable characteristics predicting post-treatment outcomes of the treated unit to those of a convex combination of untreated units, denoted as a synthetic control. Abadie et al. (2010) argue, using a linear factor model, that matching on observed confounders and on a long set of pre-treatment outcomes is possible as long as unobserved and observed confounders are also matched. It remains to be verified for any particular application, however, to which extent this matching is achieved. Importantly, the proposed linear factor model allows individual effects (factor loadings) to have time-varying effects (factors) on the outcome, which generalizes the DID approach that restricts the latter to be constant over time, as discussed in Gobillon and Magnac (2016).

In this specific application, for  $i = 1, \dots, J + 1$  municipalities and  $t = 1, \dots, T$  time periods with  $T_0$ ,  $1 \leq T_0 < T$ , pre-UPS treatment periods, let  $Y_{it}^N$  be the waste generation of municipality  $i$  in time  $t$  in the absence of the UPS policy and  $Y_{it}^{UPS}$  be its waste generation if exposed to the policy. Without loss of generality, let the first municipality,  $i = 1$ , be exposed to

the policy intervention while the remaining  $J$  municipalities are not. The policy causal effect to be estimated is given by the Treatment effect on the Treated,  $TT_t = Y_{1t}^{UPPS} - Y_{1t}^N$  for  $t > T_0$ , and the empirical challenge is to reconstruct the counterfactual  $Y_{1t}^N$ , i.e., the waste generation outcome of the treated municipality after the intervention had it not been treated.<sup>7</sup> Once the counterfactual outcome,  $\hat{Y}_{1t}^N$ , is estimated, the average causal effect of the policy over the periods after treatment, i.e., the Average Treatment effect on the Treated (ATT), is computed as  $\frac{1}{T-T_0} \sum_{t>T_0} (Y_{1t} - \hat{Y}_{1t}^N)$ .

The synthetic control method recreates this counterfactual with a convex combination of untreated municipalities, i.e.,  $\hat{Y}_{1,t>T_0}^N = \sum_{i \neq 1} w_i^* Y_{i,t>T_0}^N$ , by choosing a vector of weights  $W^* = \{w_i^*\}_{i \neq 1}$  through an optimization program. More specifically, let  $\mathbf{H} = (\eta_1, \dots, \eta_{T_0})$  be a set of weights that generates a linear combination of pre-treatment waste generation outcomes  $Y_i^H = \sum_{t=1}^{T_0} \eta_t Y_{it}$  of a municipality  $i$ , and take  $M$  of such combinations. Now, let  $X_1$  be a  $(L + M) \times 1$  predictor vector composed by all  $M$  selected linear combinations of pre-treatment outcomes and  $L$  observed waste generation determinants of the treated municipality, i.e.,  $L$  covariates. Finally, take  $X_0$  to be a matrix of dimensions  $(L + M) \times J$ , in which each column is the equivalent of the  $X_1$  vector for an untreated municipality.<sup>8</sup> Then, for the treated municipality,  $W$  is chosen to minimize  $\sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$  subject to  $w_i \geq 0$ ,  $\forall i$  untreated municipalities, and  $\sum_{i \neq 1} w_i = 1$ , where  $V$  is a  $(L + M) \times (L + M)$  diagonal weighting matrix for each predictor variable in the matrices  $X$ .

Different specifications of the estimator are possible by changing which  $M$  combinations of pre-treatment outcomes are used and how they are weighted, along with the  $L$  observed covariates, in the matrix  $V$ . In this paper, as in Abadie and Gardeazabal (2003), we select a sole combination of pre-treatment outcomes, i.e.  $M = 1$ , corresponding to their unweighted average, i.e.  $Y_i^H = \frac{1}{T_0} \sum_{t=1}^{T_0} Y_{it}$ . The predictors weighting matrix  $V$ , in turn, is chosen, also as in Abadie and Gardeazabal (2003) and Abadie et al. (2010), among positive definite and diagonal matrices such that the Mean Squared Error (MSE) of the outcome is minimized for the pre-treatment periods.

Since this comparative case study setting is not suitable for large sample

---

<sup>7</sup>Note that, as usual, it is assumed that the intervention has no anticipation effects on the waste generation before its implementation in any of the  $J+1$  cities, i.e.  $Y_{it}^{UPPS} = Y_{it}^N$ ,  $\forall i$  and  $t$ ,  $1 \leq t \leq T_0$ ; and that there are no spillovers from the intervention on waste outcomes of untreated municipalities, i.e.  $Y_{it}^{UPPS} = Y_{it}^N \forall t$  and  $i > 1$ .

<sup>8</sup>In order to avoid interpolation biases the control group is limited to municipalities with similar observed covariates and pre-treatment outcomes.



inferential techniques, placebo tests are performed by applying the SCM to each control municipality in the sample.<sup>9</sup> If the estimated treatment effect in the treated municipality is large relative to the one estimated for a control municipality chosen at random, it is possible to conclude that the UPS policy had a significant impact on waste generation in the treated municipality. Placebo tests are also performed in a restricted sample of control units for which the estimated synthetic controls provide a better pre-treatment fit, in the form of a lower MSE than the actual treated unit. Finally, the probability of finding treatment effects as high as in the treatment unit is reported as  $\frac{\sum_{j=2}^{J+1} \mathbf{1}(\frac{1}{T-T_0} \sum_{t>T_0} TT_{tj} \geq \frac{1}{T-T_0} \sum_{t>T_0} TT_{t1})}{J+1}$ .

Lastly, in order to illustrate the methodological innovation with respect to the previous UPS literature, the SCM results are compared to the ones obtained through the conventional difference-in-differences approach. For this, the same  $L$  covariates specified in the matrices  $X$  above are used to estimate the ATT. Finally, the bias of the DID estimator is evidenced by showing that the parallel trend assumption fails in the pre-treatment period.

### 3 Background and Data

We analyze the UPS policy implemented in Trento, Italy, in January 2013. Trento’s implementation of UPS consists of adding, on top of the already existing flat fee for MSW collection,<sup>10</sup> a unit price of € 0.09 per liter of unsorted waste. Therefore, UPS was introduced without changing any other ongoing baseline policy. In particular, recycling waste remains unpriced and collected at the curb. Starting from January 2013, each household was mandated to discard its unsorted waste either in 30 liter distinctively marked waste bags that cost € 2.69 each or, in buildings with a maximum of four flats and single houses, in 120 liter waste bins equipped with a coded microchip that is activated at every emptying of the bin at the price of € 10.8. To avoid free riding problems, every building’s bin is locked. If non-compliant bags are found, the whole building is charged for the offense to guarantee self-enforcing and mutual monitoring among residents (see, e.g., Reschovsky and Stone, 1994, for a discussion).

<sup>9</sup>As importantly suggested by one referee, other authors also implement bootstrap methods to estimate confidence intervals around the synthetic’s estimate (see, e.g., Corral and Schling, 2017; Sills et al., 2015; Kirkpatrick and Benneer, 2014).

<sup>10</sup>This flat fee is paid by all Italian municipalities, and depends on factors such as the floor area of the house and the number of inhabitants, and not on the actual amount of waste produced.

We collected monthly data on waste generation,  $Y$ , between January 2008 and December 2016 for nineteen Italian municipalities in different regions: Trento – the treated city –, Bari, Benevento, Bergamo, Brindisi, Chivasso, Ciriè, Collegno, Milano, Moncalieri, Novara, Pesaro, Pinerolo, Rivoli, Salerno, Settimo Torinese, Taranto, Torino, and Venaria Reale. MSW data for the region of Piedmont, Apulia, and Campania are collected from the respective regional observatory on municipal waste,<sup>11</sup> while MSW data for Trento come from the official website of the municipality<sup>12</sup>. For the other municipalities, MSW data was provided upon request either by the regional environmental protection agency (ARPA - Pesaro) or by the waste collection company (A2A Bergamo, AMSA Milano, and ASSA Novara).<sup>13</sup>

The municipalities under analysis are all those with observed waste generation determinants reasonably close to those of Trento’s, for which the authors could access detailed data.<sup>14</sup> Following one of the seminal papers on the SCM by Abadie et al. (2010), we restrict our control group to units with similar observed determinants to those of the treated unit in order to decrease possible interpolation biases. Thus, the control group includes units without UPS and with baseline policies akin to Trento including recycling programs such as curbside collection of recycling.<sup>15</sup>

The outcome variables of interest are the three waste streams measured as log per capita kilograms of unsorted waste ( $\ln UW$ ), recycling waste ( $\ln RW$ ), and total waste ( $\ln TW$ ). The amount of total waste is computed as the sum of recycling and unsorted waste, with recycling being recyclable items (e.g., glass, paper, textiles) that arrive to the recycling facility for treatment, and unsorted waste being the residual fraction that goes to landfill and incineration.<sup>16</sup> The covariates determining waste generation were chosen by literature review, and represent socio-economic predictors of waste generation and recycling habits.<sup>17</sup> These are: average household

<sup>11</sup>Respectively, [www.cittametropolitana.torino.it](http://www.cittametropolitana.torino.it), [www.sit.puglia.it](http://www.sit.puglia.it), [www.mysir.it](http://www.mysir.it).

<sup>12</sup>[www.comune.trento.it](http://www.comune.trento.it).

<sup>13</sup>The control group includes municipalities with available data that approximate Trento’s waste generation determinants including, e.g., curbside recycling.

<sup>14</sup>Since, in Italy, reporting monthly-level MSW data is not required by law, monthly time series are unavailable for most municipalities.

<sup>15</sup>Curbside collection was introduced in Trento in 2006 and in the control cities of Bergamo in 2000, Ciriè in 2005, Chivasso and Pesaro in 2006, Moncalieri and Novara in 2004, and Torino in 2003. We do not possess records of collection calendars and bin sizes, yet, these are relatively standard across cities according to anecdotal evidence provided by the Regional Environmental Protection Agency (ARPA).

<sup>16</sup>Note that in the aggregation of recycling amounts we did not correct for the seasonality of sub-streams (e.g., green waste), as this seasonality is common to all cities.

<sup>17</sup>See, e.g., Grossmann et al. (1974); Jenkins et al. (2003); Miranda and Bauer (1996);

size (*hhSize*); log per capita income in thousand euros (*lnIncome*); income inequality, as measured by the Gini coefficient (*Gini*); educational attainment, measured as the share of the population with a graduate degree or higher (*college*); tourism intensity, measured as the nights spent by tourists divided by the local population (*tourism*); and age structure, decomposed into citizens under 15 (*age<15*) and over 65 years old (*age>65*). This data was obtained from the Italian National Institute of Statistics (Istat), as well as Comuni-Italiani.it and Tuttitalia.it.

Importantly for the SCM, the values of waste generation outcomes and predictors of the treated municipality have to lie within the convex hull spanned by the control municipalities' values, such that a convex combination of control municipalities can fit the actual treated municipality. Figure 1 provides evidence for the validity of this common support assumption by showing boxplots of mean-corrected variables for treated (grey) and control units (white).

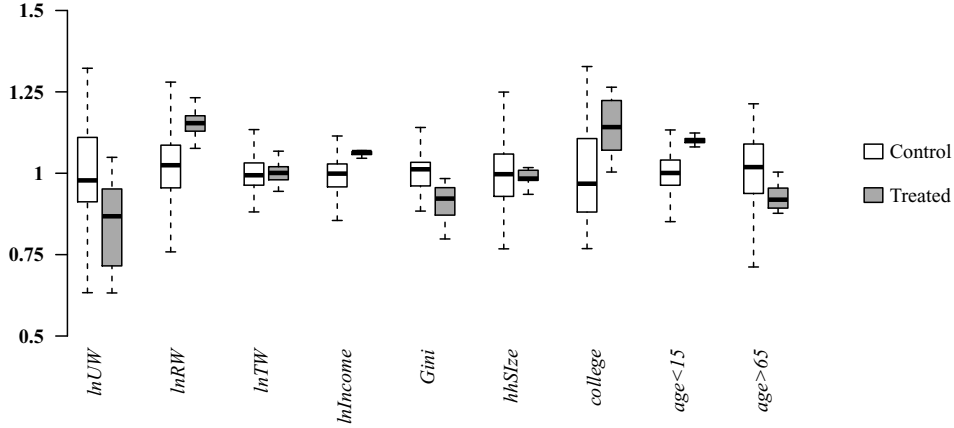
In addition, Figure 1 gives indication of possible reasons for policy endogeneity.<sup>18</sup> In terms of waste generation, Trento has on average higher (lower) levels of recycling (unsorted) waste than other control municipalities, and higher total waste, while, in terms of socio-economic characteristics, Trento has, for instance, higher average education levels and per capita income. In this respect, previous studies suggest that low unsorted waste, high recycling as well as high education and income levels are likely associated to high environmental awareness and selection bias due to an environmental activism effect (Dijkgraaf and Gradus, 2004). On the other hand, this could also point to lower opportunity costs of recycling due to, e.g, time and space. Additionally, relatively high total waste may point to selection bias due to excess overall waste and social costs.

---

Richardson and Havlicek (1978); Van Houtven and Morris (1999), and Wertz (1976).

<sup>18</sup>Note that pre-policy data for treated and control units have similar distributions to the full data plotted in Figure 1.

Figure 1: Data Descriptives: Boxplots of Mean-corrected Variables



## 4 Results

Using the SCM, we estimate the counterfactual, aka synthetic Trento, as a linear combination of outcomes of the  $J$  control municipalities. Thereby, we estimate optimal weights,  $\{w_i^*\}_{i \neq 1}$ , for each waste stream separately.

Table 1: Synthetic Trento: Estimated Weights for Control Units

| Control Units | UW    | RW    | TW    | Control Units    | UW    | RW    | TW    |
|---------------|-------|-------|-------|------------------|-------|-------|-------|
| Bari          | 0     | 0     | 0     | Novara           | 0.645 | 0.559 | 0.006 |
| Benevento     | 0.058 | 0     | 0.001 | Pinerolo         | 0     | 0     | 0     |
| Bergamo       | 0     | 0.217 | 0.536 | Pesaro           | 0.291 | 0.224 | 0.073 |
| Brindisi      | 0     | 0     | 0     | Rivoli           | 0     | 0     | 0.001 |
| Collegno      | 0     | 0     | 0.002 | Salerno          | 0     | 0     | 0     |
| Chivasso      | 0.001 | 0     | 0.363 | Settimo Torinese | 0     | 0     | 0.001 |
| Ciriè         | 0     | 0     | 0.012 | Taranto          | 0     | 0     | 0     |
| Milano        | 0     | 0     | 0     | Torino           | 0     | 0     | 0.001 |
| Moncalieri    | 0     | 0     | 0.001 | Venaria Reale    | 0.005 | 0     | 0.003 |

Table 1 reports the estimated weights. The control cities of Novara and Pesaro best resemble Trento in terms of unsorted waste (UW); Novara, Pesaro, and Bergamo in terms of recycling (RW); Bergamo and Chivasso in terms of total waste (TW). The SCM provides unbiased counterfactual estimates if predictor variables as well as outcomes of the treated unit are sufficiently close to those of the synthetic unit pre-treatment. Table 2 shows that, compared to the sample average assigning equal weights to all control

units, synthetic Trento is most similar to Trento in terms of outcomes and predictor averages, and it has a lower MSE for each waste type.

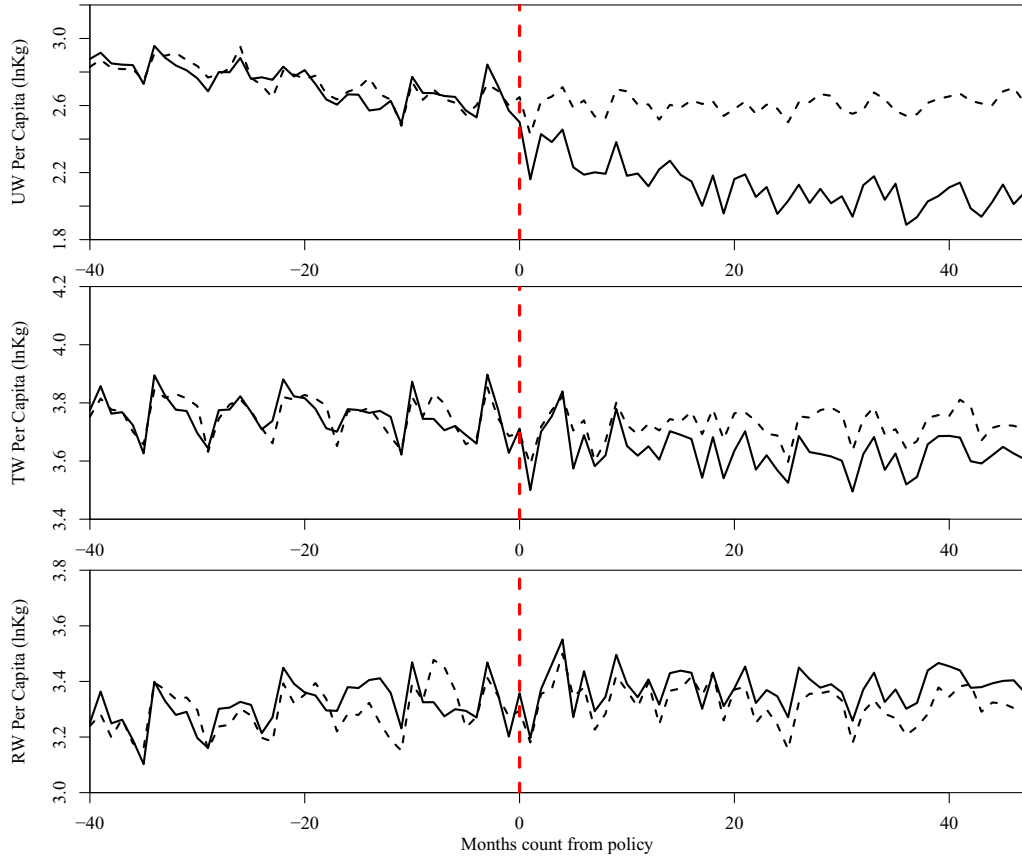
Table 2: Outcome and Predictor Means (2008-2012)

| $Y$ : $\ln UW$ | Treated | Synthetic | Sample | $Y$ : $\ln RW$ | Treated | Synthetic | Sample |
|----------------|---------|-----------|--------|----------------|---------|-----------|--------|
| $\ln UW$       | 2.735   | 2.741     | 3.049  | $\ln RW$       | 3.310   | 3.293     | 2.848  |
| $\ln RW$       | 3.270   | 3.257     | 2.811  | $\ln UW$       | 2.819   | 2.820     | 3.088  |
| $hhSize$       | 2.892   | 2.770     | 2.709  | $hhSize$       | 2.243   | 2.198     | 2.280  |
| $\ln Income$   | 2.847   | 2.734     | 2.664  | $\ln Income$   | 2.892   | 2.840     | 2.709  |
| $Gini$         | 0.274   | 0.299     | 0.309  | $Gini$         | 0.274   | 0.296     | 0.309  |
| $tourism$      | 4.736   | 4.245     | 2.237  | $tourism$      | 4.736   | 4.173     | 2.237  |
| $age < 15$     | 0.144   | 0.128     | 0.129  | $age < 15$     | 0.144   | 0.129     | 0.129  |
| $age > 65$     | 0.197   | 0.218     | 0.213  | $age > 65$     | 0.197   | 0.224     | 0.213  |
| $college$      | 0.117   | 0.115     | 0.101  | $college$      | 0.117   | 0.117     | 0.101  |
| MSE            |         | 0.029     | 0.732  | MSE            |         | 0.036     | 0.730  |

| $Y$ : $\ln TW$ | Treated | Synthetic | Sample |
|----------------|---------|-----------|--------|
| $\ln TW$       | 3.758   | 3.759     | 3.721  |
| $hhSize$       | 2.243   | 2.146     | 2.280  |
| $\ln Income$   | 2.892   | 2.876     | 2.709  |
| $Gini$         | 0.274   | 0.296     | 0.309  |
| $tourism$      | 4.736   | 3.465     | 2.237  |
| $age < 15$     | 0.144   | 0.117     | 0.129  |
| $age > 65$     | 0.197   | 0.205     | 0.213  |
| $college$      | 0.117   | 0.112     | 0.101  |
| MSE            |         | 0.203     | 0.785  |

To show the outcome fit in each time period, Figure 2 plots waste time series for Trento (solid) and synthetic Trento (dotted), and a vertical dotted line that separates pre- from post-treatment periods in time zero (January 2013).

Figure 2: UW, TW, RW Time Series for Trento (solid) and Synthetic Trento (dotted)



The behavior of the treated outcome and its counterfactual are very similar in the pre-treatment period. This suggests that counterfactuals capture treated-specific unobserved heterogeneity in waste generation, which indicates unbiasedness of the synthetic control estimators. As for the post-treatment period, Figure 2 displays the waste paths and gaps between treated and synthetic unit.<sup>19</sup>

For UW, we observe a large and persistent divergence between Trento's UW path and its counterfactual. While the latter remains fairly stable over time, the former drops immediately after policy, and levels off in the long-run. The ATT, computed as the average post-treatment percent gap, amounts to 37.5%. For TW, post-treatment gaps are negative, and of moderate magnitude at first, but then increasing in the longer-run. In particular, the treated outcome decreases to a lower level after about one year from pol-

<sup>19</sup>The sum of the counterfactual UW and RW should approximately equal the counterfactual TW. This is indeed the case, with a negligible overestimation of TW post-treatment by 0.4% on average.

icy implementation, while the counterfactual level stays constant over time. The negative ATT amounts to 8.6%. For RW, we observe mostly positive post-treatment gaps, though of very small magnitude, and increasing only after about two years from the policy. Further, both the treated outcome and its counterfactual follow a positive trend, indicating overall increasing trends in recycling habits. The ATT of 6.1% for RW, therefore, may not be such a sizable post-policy gap for the treated unit as it was for the other waste streams.

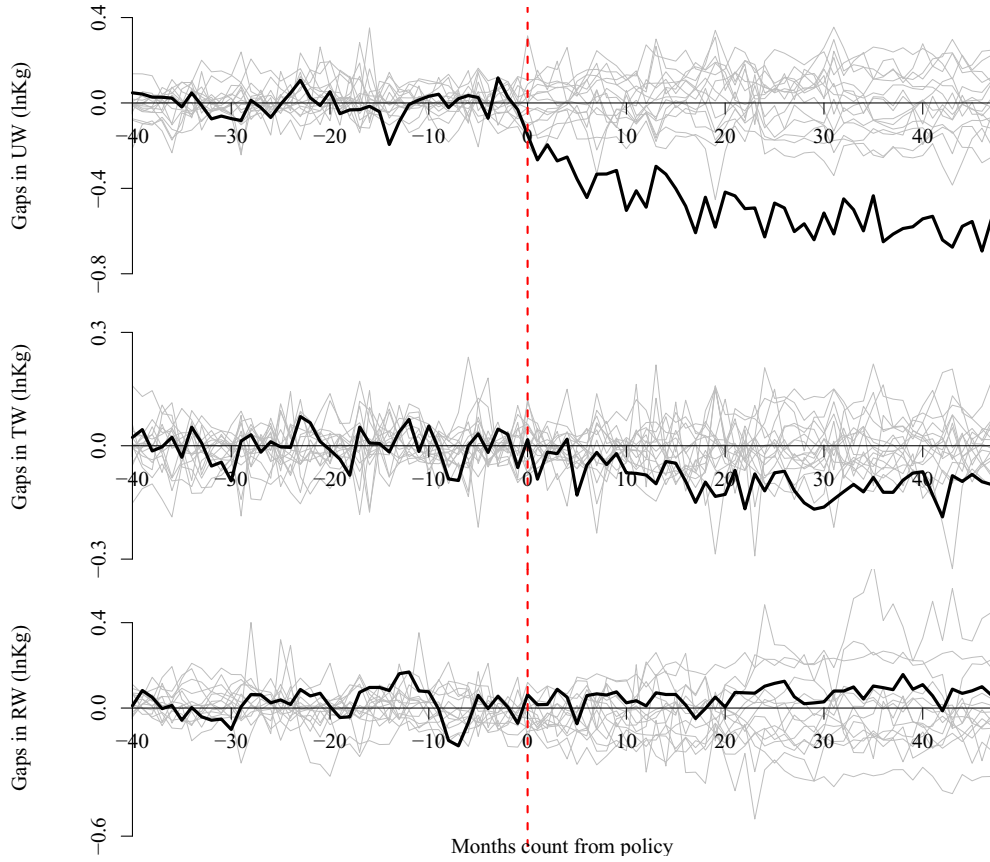
To assess the statistical significance of the gaps, we perform placebo tests by applying the synthetic control method to each control municipality in the sample.<sup>20</sup> This means to consider control units as treated one at a time, to estimate their respective synthetic control, and to compute the treatment effect given by the post-policy differences between control unit outcomes and their counterfactuals. If the estimated treatment effect for the actually treated unit, Trento, is large relative to the ones estimated for the control municipalities, the significance of the estimated effects is ascertained. Figure 3 plots gaps for the treated unit (black) and placebo control units (grey) against their respective synthetic estimate (x-axis).<sup>21</sup>

---

<sup>20</sup>All results are robust to leave-one-out estimations of the synthetic control (Abadie et al., 2015). These are available upon request.

<sup>21</sup>We exclude the treated unit from placebo counterfactual estimation because it self-selected into the policy. Additionally, counterfactuals with a pre-treatment MSE five times higher than the treated unit's MSE are discarded to safeguard against out-of-support estimations (Abadie et al., 2010).

Figure 3: Placebo Tests: Trento (black line) and Control Units (grey lines)



For UW and TW, the estimated ATT for the treated unit is the largest in absolute terms over the other 16 placebo ATT estimates, however, this does not hold for RW. Thereby, placebo tests indicate that the policy has a statistically significant impact on UW and TW, but not on RW.

A second way to account for pre-treatment goodness of fit is to compare the distribution of the ratio of post- and pre-treatment MSE for treated and placebo units (Abadie et al., 2010). Finally, since the MSE criterion overweights large discrepancies, the ATT of each unit is also plotted, for robustness, against its pre-treatment Mean relative Absolute Deviation (MAD), the pre-treatment average percent gap between treated and synthetic outcomes (Seifert and Valente, 2018).



Figure 4: MSE Ratio and ATT-MAD Tests for Trento (black) and Control Units (white)

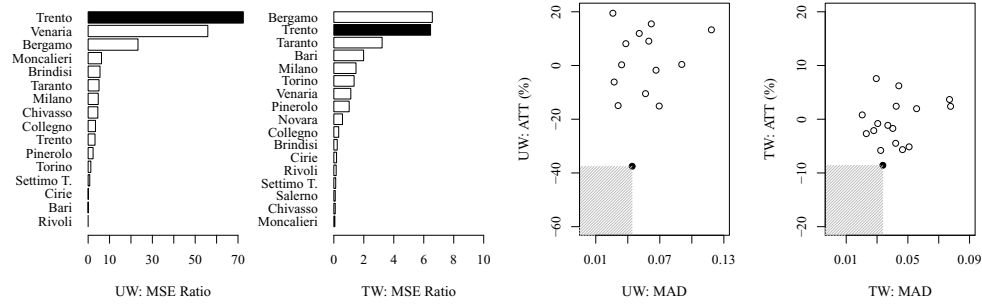


Table 4 plots the MSE ratio test and the ATT-MAD test for UW and TW. MSE ratio tests show that the treated unit has the largest, and second largest MSE ratio for UW and TW, respectively. The policy effect, hence, seems to be significant on UW and not largely on TW. Nevertheless, since no control unit shows a higher ATT as well as a smaller MAD than the treated unit, ATT-MAD tests suggest statistical significant causal effects for both UW and TW.

The statistically significant effect of the policy, stronger on UW, and weaker on TW, may be explained by changes in waste generation behaviors of households due to the policy. As intended by the policy, households decreased priced waste generation. Furthermore, reductions in TW occurring especially in the long-run suggest behavioral adjustments towards waste avoidance. This behavioral response can be also explained by increased reuse opportunities as, e.g., many supermarkets started providing facilities for refilling of beverages and detergents. On the other hand, causal effects on RW are not significant, which might be due to two reasons: First, households may further adjust their purchasing behavior by, e.g., increasing the use of lightweight packaging while also recycling more items. Because our outcome is measured in kilograms, we are not able to observe such an increase in recycling, which would lead to the underestimation (overestimation) of households' substitution (prevention) behaviors. Second, due to decreasing marginal returns to recycling efforts, further increases in recycling are less likely in municipalities with already high recycling levels such as Trento.

In conclusion, the SCM estimates suggest that the policy was effective on the priced waste, and further caused households to avoid waste generation, likely as a consequence of pro-environmental changes in waste generation behavior. However, the policy did not cause households to significantly

substitute priced waste with recycling. Concerning negative side effects of the policy, illegal dumping of waste could have happened. Yet, official statistics provided by Trento’s municipal police indicate that this was not the case.<sup>22</sup>

Finally, in order to compare our empirical strategy with the previous literature, we estimate policy effects using the conventional DID approach. The DID estimator (*ATT*) is defined as the coefficient of the interaction between a dummy for the treated unit (*treated*), i.e., a unit fixed effect equal to one for Trento in all time periods, and a time fixed effect for the post-policy periods (*Post*) equal to one only from 2013 onward. Table 3 reports estimates accounting for possible dependence in the residuals using Driscoll and Kraay (1994) adjustment.<sup>23</sup>

---

<sup>22</sup>As reported by the National Institute of Statistics and by official statistics of Trento’s municipal police, the illegal dumping of MSW follows a rather constant, linearly decreasing path, with five (twelve) cases on average in the four years after (before) policy (ISTAT, 2018; Police, 2018).

<sup>23</sup>Note that while waste outcomes are measured monthly, covariates are measured annually, which is consistent with SCM application. For DID, however, we measure all dependent and independent variables annually. Further, *hhSize* is excluded to decrease the noise generated by the linear dependence of the covariates (based on diagnostics in Fox and Monette, 1992).

Table 3: DID Regression Estimates (2008-2016) for UW, RW and TW (lnKg)

|                         | $\ln UW$                 | $\ln RW$             | $\ln TW$             |
|-------------------------|--------------------------|----------------------|----------------------|
| <i>lnIncome</i>         | -0.233<br>(0.308)        | -0.443**<br>(0.210)  | -0.416***<br>(0.099) |
| <i>Gini</i>             | -7.169***<br>(1.165)     | 2.408<br>(1.749)     | -3.638***<br>(0.194) |
| <i>tourism</i>          | 0.046***<br>(0.011)      | 0.003<br>(0.006)     | 0.029***<br>(0.006)  |
| <i>age&lt;15</i>        | 0.057***<br>(0.021)      | -0.184***<br>(0.018) | -0.020**<br>(0.008)  |
| <i>age&gt;65</i>        | 0.008<br>(0.019)         | 0.079***<br>(0.013)  | 0.028***<br>(0.006)  |
| <i>college</i>          | -0.189***<br>(0.025)     | 0.184***<br>(0.015)  | -0.034**<br>(0.016)  |
| <i>treated</i>          | -0.316***<br>(0.120)     | 0.687***<br>(0.071)  | 0.055<br>(0.040)     |
| <i>post</i>             | 0.216*<br>(0.117)        | -0.543***<br>(0.106) | -0.095**<br>(0.046)  |
| <i>ATT</i>              | -0.555***<br>(0.057)     | 0.012<br>(0.061)     | -0.076***<br>(0.024) |
| Constant                | 9.348***<br>(0.504)      | 4.583***<br>(0.594)  | 8.413***<br>(0.114)  |
| Obs.                    | 171                      | 171                  | 171                  |
| Time effects            | Yes                      | Yes                  | Yes                  |
| F Statistic             | 29.90***                 | 12.81***             | 16.56***             |
| Adjusted R <sup>2</sup> | 0.478                    | 0.464                | 0.559                |
| Notes:                  | ***p=.01; **p=.05; *p=.1 |                      |                      |

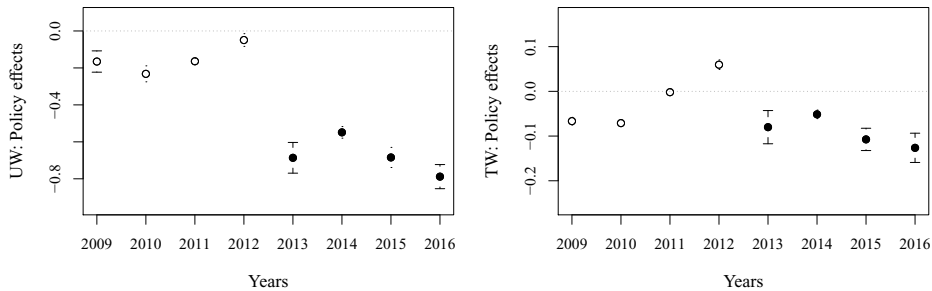
Control variable estimates present overall plausible signs and magnitudes. In particular, we find that higher income municipalities produce less total and recycling waste, with wealthier households potentially having higher opportunity costs of recycling and being overall less incline to public good provision (Magnani, 2000). In addition, the income inequality measure (*Gini*) negatively correlates with unsorted and total waste, possibly because income inequality is increasing in most cities, and at country-level, while waste streams are progressively diminishing. In addition, cities with more nights spent by tourists per inhabitant (*tourism*) produce more UW

and TW, likely because tourists have less incentives to avoid waste and recycle. We also find that older cities ( $age > 65$ ) recycle more and produce more waste in total, while cities with on average larger families ( $age < 15$ ) generates less total waste, likely because they buy in bulk; they also recycle less and produce more unsorted waste as a possible consequence of more stringent time constraints, and higher opportunity costs of recycling. Then, higher-educated cities (*college*) produce less waste and recycle more, suggesting that higher education could correlate with higher environmental awareness levels.

Concerning the estimates of the policy average causal effect, we obtain a statistically significant negative ATT of about 55% for UW, and 7.6% for TW, and, as with SCM estimates, no statistically significant ATT for RW.<sup>24</sup> For a more precise comparison of the DID and SCM estimates, we compute annual averages of the synthetic control monthly effects. After doing so, ATTs obtained by DID can be compared to SCM estimates that amount to  $-27\%$ ,  $4.2\%$ ,  $-4.1\%$  for UW, RW, and TW, respectively. Since the two estimators take substantially different approaches, this comparison is the most informative and accurate possible.

However, as previously discussed, the DID estimation is likely biased due to a violation of its Parallel Trend Assumption (PTA). We show this by performing in-time placebo tests for UW and TW (Autor, 2003). Figure 5 plots the estimated ATTs post-treatment (black) and placebo pre-treatment (white) for each year with their respective confidence interval.

Figure 5: In-time Placebos: Pre- (white) and Post-treatment (black) Policy Effects



For UW and TW, pre-treatment effects are jointly statistically significant at the 1% and 10% level, respectively, indicating a failure of the PTA

<sup>24</sup>Since the variables included in the DID estimation are correctly specified in annual terms, this is not an issue for the internal validity of our estimation.

assumed in the DID regression model. Specifically, for UW, we find statistically significant, negative effects in each pre-policy year. This suggests that UW generation trend in the treated unit prior to the policy was diverging from that of the DID counterfactual, leading to an overestimation of the UPS’ treatment effect. Concerning TW, both negative and positive statistically significant effects are estimated already pre-policy, suggesting diverging trends, and the bias of the DID estimator.

As discussed, DID only allows for individual effects that have time-invariant effects on the outcome, which might not be the case in our setting. If, e.g., environmental-friendly attitudes or time and space availability have evolving effects over time as outlined earlier, then unobserved determinants will have varying effects in time. Controlling for this variation is at the heart of this paper’s motivation to employ the SCM.

## 5 Conclusions

This paper evaluates the effects of a Unit Pricing System (UPS) implemented in Trento, Italy, in 2013, on the disposal of municipal solid waste. We use a unique data on monthly waste generation for nineteen municipalities in Italy over the period 2008-2016. To account for time-varying effects of unobserved determinants of waste generation, we employ the synthetic control method (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010), in which the counterfactual outcome of the treated municipality is reconstructed with a convex combination of untreated municipalities with similar outcome trajectories but not selecting into treatment.

We find that UPS was effective, and caused changes in household waste generation behavior. In particular, households responded by reducing priced waste, i.e. unsorted waste, by 37.5% on average in the post-treatment period. This partial causal effect seems to have been driven by waste avoidance and, possibly, by increased recycling. In fact, our results show that total waste decreases by 8.6% on average after policy, and recycling increases by 6.1%. The statistical significance of the results, found to be stronger for unsorted waste, and weaker for total waste, may be explained by changes in waste generation behaviors of households after policy. As intended, UPS induced households to decrease the generation of the priced waste and sustain increasing recycling levels. However, the policy did not cause households to

significantly substitute priced waste by recycling. In this respect, our results are in line with the literature that finds evidence for sizable unsorted waste reductions and relatively smaller, often insignificant increases of recycling (see, e.g., Allers and Hoeben, 2010; Dahlén and Lagerkvist, 2010; Fullerton and Kinnaman, 2000; Jenkins et al., 2003).

Our findings contribute to the policy debate on policy instruments for municipal waste reduction. We find strong evidence that monetary incentives are effective to induce behavioral changes in household waste generation. The results show that unit pricing systems may help to apply the European Union’s priority of waste reduction (EU, 2008), and to partly internalize negative externalities of waste generation.

Moreover, municipalities might support UPS’ effectiveness in several ways, given that rational households adjust their behavior such that their marginal cost of waste reduction equals the unit price (Palmer et al., 1997). For example, municipalities may consider decreasing costs of policy compliance by, for instance, facilitating curbside recycling and awareness-raising campaigns that make recycling more appealing, i.e., less time and space consuming. Yet, introducing UPS in municipalities with high recycling may have a limited impact because households sort most of the goods in their bundle already before policy. For this reason, municipalities with lower recycling than Trento may experience stronger UPS’ effects on this type of waste.

Policy makers should also take into account municipalities’ heterogeneity. In fact, opportunity costs of waste avoidance, recycling, and illegal dumping likely depend on municipal characteristics such as income and education (Callan and Thomas, 2006). For example, higher-income municipalities may value their time more greatly leading to higher opportunity costs of policy compliance. On the other hand, these municipalities may face lower opportunity costs due to higher education and environmental awareness levels. Further, law enforcement, civic mindedness and social norms likely determine opportunity costs of illegal dumping. To increase the efficiency of UPS, municipalities should reduce illegal disposal opportunities, and possibly adjust the unit price accounting for households’ opportunity costs. In this respect, future research would benefit from analyzing the optimality of the unit price, and how municipal heterogeneity impacts policy effects.

Moreover, future work could also investigate the amount of social and,

especially, external cost savings from waste avoidance. The latter is indeed the most desirable outcome from a social viewpoint because, differently from recycling, it causes no private and external costs of, e.g., collection and treatment. Finally, studies are needed for the estimation of private savings due to waste reductions induced by UPS.

## Acknowledgements

We thank Manuel Arellano, Sylvain Chabé-Ferret, Jacint Enrich, Bernd Fitzenberger, Yinghua He, Thierry Magnac, Roberto Renó, and Stefan Seifert for fruitful discussion and input. We also thank the editor and three anonymous referees that helped to improve the quality of the paper. We also thank valuable comments received at presentations held at the Italian Association of Environmental and Resource Economists (IAERE, Rome), Econometrics of Panel Data and Network Analysis Summer School (CRC, Berlin), European University Institute (FSR, Florence), the Italian Workshop on Econometrics and Empirical Economics (IWEEE, Milan), the 6<sup>th</sup> World Congress of Environmental and Resource Economists (WCERE, Gothenburg), and the Counterfactual Methods for Policy Impact Evaluation conference (COMPIE, Berlin). Matheus Bueno acknowledges the support by the H2020-MSCA-RISE project GEMCLIME-2020 GA No. 681228 with gratitude.

## References

- Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association* 105(490), 493–505.
- Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative politics and the synthetic control method. *American Journal of Political Science* 59(2), 495–510.
- Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review* 93(1), 113–132.
- Abbott, A., S. Nandeibam, and L. O’Shea (2013). Recycling: Social norms and warm-glow revisited. *Ecological Economics* 90, 10–18.
- Acemoglu, D., S. Johnson, A. Kermani, J. Kwak, and T. Mitton (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics* 121(2), 368–391.

- Alesina, A. F., G. Tabellini, and F. Trebbi (2017). Is Europe an optimal political area? *NBER Working Paper No. w23325*.
- Allers, M. and C. Hoeben (2010). Effects of unit-based garbage pricing: A differences-in-differences approach. *Environmental Resource Economics* 45(3), 405–28.
- Almer, C. and R. Winkler (2017). Analyzing the effectiveness of international environmental policies: The case of the Kyoto Protocol. *Journal of Environmental Economics and Management* 82(C), 125–151.
- Autor, D. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21(1), 1–42.
- Bel, G. and R. Gradus (2016). Effects of unit-based pricing on the waste collection demand: A meta-regression analysis. *Resource and Energy Economics* 44, 169–182.
- Bénabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies* 70(3), 489–520.
- Besley, T. and M. Reynal-Querol (2011). Do democracies select more educated leaders? *American Political Science Review* 105(3), 552–566.
- Billmeier, A. and T. Nannicini (2013). Assessing economic liberalization episodes: A synthetic control approach. *The Review of Economics and Statistics* 95(3), 983–1001.
- Brekke, K., S. Kverndokk, and K. Nyborg (2004). An economic model of moral motivation. *Journal of Public Economics* 87(9-10), 1967–1983.
- Bruvold, A. and K. Nyborg (2003). The cold shiver of not giving enough: On the social cost of recycling campaigns. *Land Economics* 80(4), 539–549.
- Bucciol, A., N. Montinari, and M. Piovesan (2015). Do not trash the incentive! Monetary incentives and waste sorting. *The Scandinavian Journal of Economics* 117(4), 1204–1229.
- Callan, S. J. and J. M. Thomas (2006). Analyzing demand for disposal and recycling services: A systems approach. *Eastern Economic Journal* 32(2), 221–240.
- Carattini, S., A. Baranzini, and R. Lalive (2018). Is taxing waste a waste of time? Evidence from a Supreme Court decision. *Ecological Economics* 148, 131–151.
- Cavallo, E., S. Galiani, I. Noy, and J. Pantano (2013). Catastrophic natural disasters and economic growth. *The Review of Economics and Statistics* 95(5), 1549–1561.
- ComuneTrento (2012). Progetto di riduzione dei rifiuti nella grande distribuzione organizzata - Dal progetto Spesa Leggera a Ecoacquisti Trentino. *Municipality of Trento*. Available at <http://www.comune.trento.it/Aree-tematiche/Ambiente-e-territorio/Rifiuti-urbani/Gestione-integrata-rifiuti/Riduzione-dei-rifiuti>.



- Corral, R. L. and M. Schling (2017). The impact of shoreline stabilization on economic growth in small island developing states. *Journal of Environmental Economics and Management* 86(C), 210–228.
- Dahlén, L. and A. Lagerkvist (2010). Pay as you throw: Strengths and weaknesses of weight-based billing in household waste collection systems in sweden. *Waste Management* 30, 23–31.
- Dijkgraaf, E. and R. H. J. M. Gradus (2004). Cost savings in unit-based pricing of household waste: The case of the Netherlands. *Resource and Energy Economics* 26(4), 353–371.
- Dijkgraaf, E. and R. H. J. M. Gradus (2009). Environmental activism and dynamics of unit-based pricing systems. *Resource and Energy Economics* 31(1), 13–23.
- Driscoll, J. C. and A. C. Kraay (1994). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80(4), 549–560.
- EEA (2009). Diverting waste from landfill. Effectiveness of waste management policies in the European Union. *European Environment Agency (EEA), Report No 7/2009*.
- EEA (2013). Managing municipal solid waste - A review of achievements in 32 European countries. *European Environment Agency (EEA), Report No 2/2013*.
- EU (2008). Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain directives. *Official Journal of the European Union* 22/11/2008, 3–30.
- European Commission (2010). Being wise with waste: The EU’s approach to waste management. *Luxembourg: Publications Office of the European Union*.
- Fox, J. and G. Monette (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association* 87(417), 178–183.
- Fullerton, D. and C. T. Kinnaman (1996). Household responses to pricing garbage by the bag. *The American Economic Review* 86(4), 971–984.
- Fullerton, D. and C. T. Kinnaman (2000). Garbage and recycling with endogenous local policy. *Journal of Urban Economics* 48(3), 419–442.
- Gobillon, L. and T. Magnac (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *The Review of Economics and Statistics* 98(3), 535–551.
- Grossmann, D., J. F. Hudson, and D. Marks (1974). Waste generation models for solid waste collection. *Journal of the Environmental Engineering Division* 100(6), 1219–1230.
- Hong, S. and R. M. Adams (1999). Household responses to price incentives for recycling: Some further evidence. *Land Economics* 75(4), 505–514.

- Hong, S., R. M. Adams, and H. A. Love (1993). An economic analysis of household recycling of solid wastes: The case of Portland, Oregon. *Journal of Environmental Economics and Management* 25(2), 136–46.
- Huang, J.-C., J. M. Halstead, and S. B. Saunders (2011). Managing municipal solid waste with unit-based pricing: Policy effects and responsiveness to pricing. *Land Economics* 87(4), 645–660.
- ISTAT (2011). Verde urbano. *ISTAT - Italian National Statistical Institute*. Available at <https://www.istat.it/it/archivio/86880>.
- ISTAT (2018). Data on environmental crimes. *Italian National Institute of Statistics - ISTAT*. Data available at <https://www.istat.it/it/archivio/218648>.
- Jenkins, R. R., S. A. Martinez, K. Palmer, and M. J. Podolsky (2003). The determinants of household recycling: A material-specific analysis of recycling program features and unit pricing. *Journal of Environmental Economics and Management* 45(2), 294–318.
- Kahn, M. (2007). Do greens drive hummers or hybrids? Environmental ideology as a determinant of consumer choice. *Journal of Environmental Economics and Management* 54(2), 129–145.
- Kim, G. S., Y. J. Chang, and D. Kelleher (2008). Unit pricing of municipal solid waste and illegal dumping: An empirical analysis of Korean experience. *Environmental Economics and Policy Studies* 9(3), 167–176.
- Kipperberg, G. (2007). A Comparison of Household Recycling Behaviors in Norway and the United States. *European Association of Environmental and Resource Economists* 36(2), 215–235.
- Kirkpatrick, J. A. and L. S. Benneer (2014). Promoting clean energy investment: An empirical analysis of property assessed clean energy. *Journal of Environmental Economics and Management* 68(2), 357–375.
- La Repubblica (2006). Istat, Trento la città più ecologica. In Italia cresce la raccolta differenziata. *La Repubblica*. Available at <http://www.repubblica.it/2007/08/sezioni/ambiente/istat-ambiente/istat-citta/istat-citta.html>.
- Linderhof, V., P. Kooreman, M. Allers, and D. Wiersma (2001). Weight-based pricing in the collection of household waste: The Oostzaan case. *Resource and Energy Economics* 23(4), 359–371.
- Magnani, E. (2000). The Environmental Kuznets Curve, environmental protection policy and income distribution. *Ecological Economics* 32(3), 431–443.
- Miranda, M. L. and S. Bauer (1996). The urban performance of unit pricing: An analysis of variable rates for residential garbage collection in urban areas. *The U.S. Environmental Protection Agency*. Report prepared for the Office of Policy, Planning, and Evaluation.

- Morris, G. E. and D. M. Holthausen Jr. (1994). The economics of household solid waste generation and disposal. *Journal of Environmental Economics and Management* 26(3), 215–234.
- O’Leary, E., D. Cunningham, and T. Coakley (2017). A review of current priorities and emerging issues in European waste policy. *Environmental Protection Agency (EPA) Report of the EPA Research Programme 2014-2020*.
- Palmer, K., H. Sigman, and M. Walls (1997). The cost of reducing municipal solid waste. *Journal of Environmental Economics and Management* 33(2), 128–150.
- Pigou, A. C. (1932). Divergences between marginal social net product and marginal private net product. *The Economics of Welfare Part II, Chapter IX*.
- Pinotti, P. (2015). The economic costs of organized crime: Evidence from southern Italy. *The Economic Journal* 125(586), 203–232.
- Podolsky, M. J. and D. V. Nestor (1998). Assessing incentive based environmental policies for reducing household waste disposal. *Contemporary Economic Policy* 16(4), 27–39.
- Police (2018). Data on the illegal dumping of municipal solid waste (provided upon request). *Municipal Police of Trento*.
- Reichenbach, J. (2008). Status and prospects of pay-as-you-throw in Europe: A review of pilot research and implementation studies. *Waste Management* 12(28), 2809–2814.
- Reimer, M. N. and A. C. Haynie (2018). Mechanisms matter for evaluating the economic impacts of marine reserves. *Journal of Environmental Economics and Management* 88(C), 427–446.
- Reschovsky, J. D. and S. E. Stone (1994). Market incentives to encourage household waste recycling: Paying for what you throw away. *Journal Of Policy Analysis and Management* 13(1), 120–39.
- Richardson, R. A. and J. H. Havlicek (1978). Economic analysis of the composition of household solid wastes. *Journal of Environmental Economics and Management* 5(1), 103–111.
- Sakai, S., T. Ikematsu, Y. Hirai, and H. Yoshida (2008). Unit-charging programs for municipal solid waste in japan. *Waste Management* 28(12), 2815–2825.
- Seifert, S. and M. Valente (2018). An offer that you can’t refuse? Agrimafias and migrant labor on vineyards in southern Italy. *DIW Berlin Discussion Papers* 1735.
- Sills, E. O., D. Herrera, and P. A. Kirkpatrick A. J. and, ... (2015). Estimating the impacts of local policy innovation: The synthetic control method applied to tropical deforestation. *PLOS ONE* 10(7), 1–15.
- Skumatz, L. (2008). Pay as you throw in the U.S.: Implementation, impacts, and experience. *Waste Management* 28(12), 2778–2785.

- Thøgersen, J. (2006). Norms for environmentally responsible behaviour: An extended taxonomy. *Environmental Psychology* 26(4), 247–261.
- Usui, T. (2008). Estimating the effect of unit-based pricing in the presence of sample selection bias under Japanese recycling law. *Ecological Economics* 66, 281–288.
- Usui, T. (2009). Does a rebound effect exist in solid waste management? *Critical Issues in Environmental Taxation* 7, 413–422.
- Usui, T. and K. Takeuchi (2014). Evaluating unit-based pricing of residential solid waste: A panel data analysis. *Environmental and Resource Economics* 58(2), 245–271.
- Van Houtven, G. L. and G. E. Morris (1999). Household behavior under alternative pay-as-you-throw systems for solid waste disposal. *Land Economics* 75(4), 515–537.
- Wertz, K. L. (1976). Economic factors influencing households’ production of refuse. *Journal of Environmental Economics and Management* 2(4), 263–272.
- Wright, C., J. M. Halstead, and J.-C. Huang (2018). Estimating treatment effects of unit-based pricing of household solid waste disposal. *Agricultural and Resource Economics Review*, 1–23.
- Yamakawa, H., K. Ueta, and Y. Terashima (2002). Factors influencing illegal dumping in communities with variable rate programs. *Journal of the Japan Society of Waste Management Experts* 13(4), 419–427.

# Heterogeneous effects of waste pricing policies

*Job market paper*

**Abstract** Using machine learning methods in a quasi-experimental setting, I study the heterogeneous effects of introducing waste prices—unit prices on household unsorted waste disposal—on waste demands and social welfare. First, using a unique panel of Italian municipalities with large variation in prices and observables, I show that waste demands are nonlinear. I find evidence of nudge effects at low prices, and increasing elasticities at high prices driven by income effects and waste habits before policy. Second, I estimate policy impacts on pollution and municipal management costs, and compute the overall social cost savings for each municipality. Social welfare effects become positive for most municipalities after three years of adoption, when waste prices cause significant waste avoidance.

**Keywords:** Waste pricing, Causal effect heterogeneity, Welfare analysis, Machine learning

# Introduction

Waste management generates pollution externalities that are not internalized by households under traditional flat pricing schemes (Fullerton & Kinnaman, 1996). To correct this inefficiency, a growing number of municipalities have implemented Pigou prices (1932) known as Pay-As-You-Throw (PAYT) programs that require households to pay for each unit (per bag, can, or weight) of unsorted waste presented for collection. Theory predicts households will substitute unsorted waste with increased recycling and/or waste avoidance (Smith, 1972).<sup>1</sup> However, empirical estimates of average price effects provide mixed evidence on the magnitude of this reduction as well as the behavioral mechanisms behind it (Bueno & Valente, 2019). Moreover, policy impacts may vary across municipalities depending, e.g., on the adopted price level and household characteristics (Kinnaman, 2006). Disentangling sources of effect heterogeneity is important to tailor effective policies and deliver insights on why some municipalities refrain from PAYT adoption.<sup>2</sup>

In this paper, I examine heterogeneous demand responses to waste prices, and their impact on social welfare. The main challenge in the analysis is that determinants of waste generation and policy adoption are possibly many, and may confound the estimation of causal effects.<sup>3</sup> This motivates the collection of a unique panel of municipalities with a large variation in prices and observables, and the estimation of municipal level causal effects of prices (continuous treatment) via machine learning methods. These techniques, in fact, allow to control for a high-dimensional set of covariates.

I estimate social welfare effects for each municipality by combining price effects on unsorted and recycling waste with their impacts on pollution and municipal costs. First, I find that price levels matter. Higher prices cause larger unsorted waste reductions and recycling increases. Lower prices cause relatively more waste avoidance. Second, municipal characteristics also matter.

---

<sup>1</sup>Waste avoidance is defined as using household effort to reduce total waste (via e.g. product reuse).

<sup>2</sup>In Italy, less than 8% of municipalities adopt PAYT (ISPRA, 2019). Policymakers fear cost increases due to, e.g., higher recycling (Facchini, 2020). For anecdotal evidence, see e.g. Gilli *et al.* (2018b), Allers & Hoeben (2010), Callan & Thomas (1999).

<sup>3</sup> Previous studies suggest that waste generation determinants may also drive policy adoption decisions. These are, e.g., households' income, education and pre-policy waste levels (see e.g. Gradus *et al.*, 2019).

While elasticities are rather constant at low prices, suggesting nudge effects, elasticities increase at high prices especially for municipalities with low income and little recycling before policy. Third, waste avoidance matters for welfare, and drives benefits for most municipalities after three years of adoption.

This paper contributes to three distinct literatures. First, my work connects to a large and growing literature on waste prices. One set of papers estimates the price elasticity of waste demands or the causal effect of PAYT (binary treatment) on waste amounts.<sup>4</sup> Compared to my study, these papers require more restrictive identifying assumptions, e.g., constant price effects across units (effect homogeneity) or deliver less precise estimates, e.g., because they exploit only one specific price change. In Bueno & Valente (2019), we have shown that time-varying effects of unobservables on waste generation lead to bias of the difference-in-differences estimator. The main difference between this paper and past studies is the analysis of heterogeneous policy effects and social welfare, while accounting for a wide range of prices, covariates, and lagged waste outcomes. Moreover, the welfare effects of PAYT policies are largely unexplored. Most relatedly, Dijkgraaf & Gradus (2015) show that waste prices are more cost-effective than other institutional modes of collection.

Second, this paper contributes to studies in behavioral environmental economics, analyzing drivers of waste behaviors.<sup>5</sup> Building on this literature, I rationalize welfare effect heterogeneity by showing that household reactions on recycling and waste avoidance are ambiguous from a theoretical viewpoint. My work also relates to the literature on nudging. Nudges, also in the form of low prices, can correct externalities by reducing, e.g., plastic bag consumption (Rivers, 2017). To my knowledge, my results provide the first empirical evidence of agents' nudgeability to waste prices, and of its social welfare impacts.

Third, to my knowledge, this is the first empirical application of machine learning for causal inference combining continuous treatment, staggered adoption, and self-selection. In addition, while random forests have been successfully applied in labor economics,<sup>6</sup> no study applied this algorithm to high-

---

<sup>4</sup>This includes, e.g., Bueno & Valente 2019; Carattini *et al.* 2018; Bucciol *et al.* 2015; Huang *et al.* 2011; Allers & Hoeben 2010; Fullerton & Kinnaman 2000, 1996. See Kinnaman 2014 for a review.

<sup>5</sup>See, e.g., D'Amato *et al.* 2016, Bowles & Polania-Reyes 2012, and Gilli *et al.* 2018b for a review.

<sup>6</sup>See, e.g., Gulyas & Pytka 2019; Athey & Wager 2019; Davis & Heller 2017.

dimensional problems in environmental economics.

I study waste generation behaviors of about 3,600 Italian municipalities over 2010-2015. Italy provides an ideal setting to study heterogeneous effects of waste prices because both price levels and socio-economic characteristics largely vary across municipalities. I use web scraped and administrative data to construct a new and rich dataset on waste generation and price adoption at the municipal level. The final dataset includes 45 different price levels ranging from 1 to 18 euro (€) cents per liter of unsorted waste, and 90 municipal characteristics that may explain price adoption and waste generation.

To consistently estimate municipal level parameters with high-dimensional data, I use machine learning-inspired matching estimators called generalized Random Forests (RFs) (Athey *et al.*, 2019).<sup>7</sup> Intuitively, RFs partition the large covariate space into small neighborhoods of municipalities mostly similar in those characteristics that drive parameter heterogeneity. Within neighborhoods, I estimate constant treatment effects by the residual-on-residual regression estimator, or R-learner (Nie & Wager, 2019).<sup>8</sup>

The advantage of using RFs is to relax the assumption of constant price effects across municipalities and estimate the full effect distribution with pointwise-consistent confidence intervals. Estimation avoids ad hoc modeling choices, and flexibly accounts for parameter heterogeneity in the large set of (often correlated) covariates.<sup>9</sup> The improvement of RFs vis-à-vis, for instance, theory-informed heterogeneity analysis is to provide with a data-driven documentation of heterogeneous causal effects, as opposed to specification search.

Using these machine learning methods, I estimate municipal level price elasticities of demands for unsorted, recycling, and total waste per capita. The hypothesis of no heterogeneity is rejected for all outcomes. Waste prices cause large unsorted waste reductions driven by increased recycling and, to a smaller extent, waste avoidance. To disentangle sources of heterogeneity, I regress these elasticities on price levels and a parsimonious set of relevant regressors captur-

---

<sup>7</sup>RFs build upon Athey & Wager (2018), Athey & Imbens (2016) and, originally, Breiman (2001).

<sup>8</sup>R-learners are robust to confounding affecting outcome and treatment (Chernozhukov *et al.*, 2017).

<sup>9</sup>Differently, standard regression methods are justified if treatment effects are constant, observables have linear or pre-specified effects, and unobservables are time-invariant (Wager, 2020).



ing household costs of waste disposal: income, education, and pre-policy waste levels.<sup>10</sup> I find a nonlinear relationship between elasticities and price levels: while at high prices (above 9 cents) elasticities are increasing, elasticities are rather constant at low prices. The estimated variation at low prices is 4% of the variation at high prices. In this range, a one cent price increase reduces unsorted waste by 5 to 10%, increases recycling by 2 to 6%, and reduces total waste by 0.1 to 0.7%. Low prices reduce total waste by more (0.6-0.8%) and increase recycling by less (2.5-3.2%).

I find no evidence of income effects at low prices. This finding suggests that low prices work as nudges, i.e. instruments that influence behavior without budgetary incidence (Farhi & Gabaix, 2020). To provide an intuition, households reducing unsorted waste by one standard trash bag (30 liters) save about 90 cents in the first price quartile (3 cents), while 4€ in the third price quartile (13 cents). As shown for plastic bag consumption (see, e.g., Rivers, 2017), low prices are symbolic but serve to remind households of the costs of waste as well as to promote recycling and avoidance behaviors.

Moreover, I find no evidence of waste prices being regressive at high prices: I estimate that high-income municipalities are less elastic, and pay more under PAYT.<sup>11</sup> This implies that municipalities could increase waste prices without distributional concerns.

Having established heterogeneity of price effects, the second part of the empirical analysis focuses on municipal waste management costs. I estimate that PAYT leaves unit costs of waste mostly unaffected, suggesting constant returns to scale.<sup>12</sup> Next, I simulate the impact of PAYT adoption on social welfare using prior estimates for the relative environmental costs of unsorted versus recycling waste (Kinnaman *et al.*, 2014). I find that waste prices can raise social costs, especially when households respond by increasing recycling only. However, after three years of adoption, I predict welfare benefits for most municipalities of on average €30 up to €170 per person. As unit costs of unsorted waste are higher than those of recycling, waste avoidance triggers

---

<sup>10</sup>Their relevance is discussed in, e.g., Bueno & Valente (2019). See Gilli *et al.* (2018b) for a review.

<sup>11</sup>Policy costs for rich vs. poor municipalities amount to on average €105 vs. €87 per capita/year. Rich municipalities are defined by having an annual per capita income above the third quartile of €16k.

<sup>12</sup>This is consistent with results on Italian municipalities by, e.g., Abrate *et al.* (2014).

large welfare benefits. This implies that low prices may be preferable from a social cost perspective although they cause comparatively smaller unsorted waste reductions.

I present additional analyses in support of the identifying assumptions of unconfoundedness and no spillover effects (no waste tourism). Results are also robust to confounding from adoption of weight versus volume systems (see Bel & Gradus, 2016, for a review of past findings). Finally, I contrast my average estimates to the binary treatment case, event-study-like difference-in-differences, and R-learning LASSO regression. Results highlight the importance to account for continuous rather than binary treatment, the bias of difference-in-differences due to violation of its identifying parallel trend assumption, and robustness to the specific choice of R-learning estimator.

The remainder of the paper is structured as follows. Section 1 describes policy background and data. Section 2 discusses the theoretical framework and empirical methodology. Section 3 presents the main results and their policy implications. Section 4 concludes.

## 1 Background and Data

### 1.1 PAYT policies

PAYT policies in Italy, as in many municipalities worldwide, require households to pay a price per unit of unsorted waste according to either its volume (per bag or bin) or weight (per kilogram). PAYT fulfills the equivalence principle for which waste service consumers pay for its consumption (as, e.g., for energy and water), and the polluter-pays principle for which households pay according to their unsorted waste.

The baseline policy in both PAYT and non-PAYT municipalities is a flat fee independent of waste quantities, namely, the unit price is zero. Flat fees depend on house ( $m^2$ ) and household size (number of inhabitants). PAYT municipalities reduce the flat fee to cover only fixed costs of waste management, and implement waste prices to cover variable costs.

Municipalities can decide whether and when to implement PAYT, as well as the price level and collection system. Policy adoption decisions are based on, for instance, goals of waste pollution and management cost reduction. Price

levels are set based on lagged and expected levels of waste generation and management costs. Yet, political and socio-economic factors may also matter (see Gradus *et al.* 2019 and Section 1.4 for details). System choices are also made at the municipal level, and depend on, e.g., demographic, geographic, and cost factors (see Appendix D.4.1 for details). Households can pay either *ex ante* via prepaid bags or *ex post* via identification (tag on bags, chip on bins, electronic keys). Generally, municipalities prefer volume over weight systems because this requires cheaper technology (Kinnaman, 2006). Yet, systems may create asymmetric incentives.<sup>13</sup> In sum, drivers of price adoption, waste generation and collection mode often overlap, and are possibly many. Their relevance is, therefore, an empirical question.

Credible enforcement and monitoring systems are crucial for policy success.<sup>14</sup> Illegal dumping is one of the main possible adverse effects. However, after about fifty years of PAYT experiences worldwide, adverse effects seem a bigger fear than reality.<sup>15</sup> In Italy, anecdotal evidence suggests that (i) waste tourism in surrounding municipalities is a rare and short-lived phenomenon, (ii) enforcement and monitoring systems allow to actually decrease illegal dumping episodes, and (iii) waste haulers encourage adoption and deem waste prices as successful (Legambiente, 2017).

## 1.2 Data

Using web scraped and administrative data, I construct a new municipal level dataset with information on waste prices, waste amounts (main outcomes) and management costs, as well as socio-economic, geographic, and political determinants of waste generation and price adoption. The final database is a panel of Northern and Central Italian municipalities over the sample period 2010-2015. I exclude municipalities with missing values as well as South and Insular

---

<sup>13</sup>For instance, volume systems can encourage waste compacting and be, therefore, less effective.

<sup>14</sup>E.g., trash bins are locked, drones and photo traps track illegal dumping, and reciprocal monitoring is enforced by charging all households in a building for lack of policy compliance (CONSEA, 2019).

<sup>15</sup>See, e.g., Bueno & Valente (2019) for Italy and Skumatz (2008) for the US. The latter found illegal dumping to be a short-term issue which lasts three months or less and involves 3% of municipalities.

Italy because arguably not suitable for being in the control group.<sup>16</sup> Treated municipalities for which second-order lags are missing are also excluded.<sup>17</sup>

The resulting municipalities are 3,574. Waste prices, the treatment, cover 1.7 million people living in 194 municipalities, of which 106 in the North-West, 82 in the North-East, and six in the Center. Most of the treated municipalities implement PAYT for the first time in 2013 (77), while the others in 2012 (48), 2014 (36), and 2015 (33). Figure 1 shows the distribution of PAYT and non-PAYT municipalities in the sample.

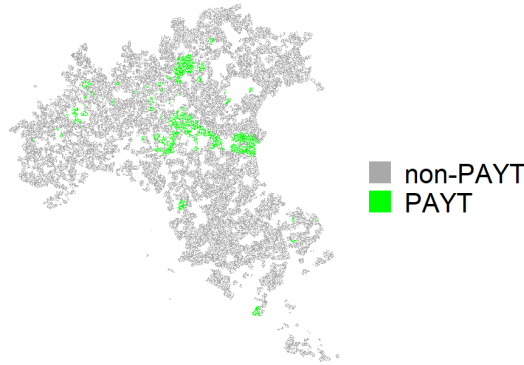


Figure 1: Map of PAYT and non-PAYT municipalities in the sample. White areas show the excluded South and Islands, municipalities with missing values, and those adopting PAYT before 2012.

This data allow for (counterfactual) predictions of policy and welfare effects for 26 million people (45% of the population) living in 3,380 municipalities without PAYT.

Variables come from a variety of sources. Waste generation and management costs are from the National Environmental Protection Agency (ISPRA). Municipal socio-economic attributes are from the National Institute of Statistics (ISTAT) and web scraping the online database of [comuni-italiani.it](http://comuni-italiani.it). Political variables on, e.g., municipal elections and mayors' characteristics are obtained from the Ministry of the Interior upon request. Geographic variables measur-

<sup>16</sup>Missing values are due to merging administrations and data errors. See Appendix B for details. In South and Insular Italy, as defined by the NUTS 1 classification, there are no treated municipalities over the sample period, and data is incline to significant mismeasurement due to illegal disposal (ISTAT, 2018).

<sup>17</sup>Policy adoption decisions correlate with second- or earlier-order outcome lags. First-order lags are biased predictors due to anticipation effects. See Section 3.4.1 for details and estimates.

ing, e.g., the distance to waste treatment sites are geocoded using the software R and data of the European Pollution Release and Transfer Register (E-PRTR).

At the time of writing, there is no open-source database on PAYT prices in Italy. I acquired the list of PAYT municipalities from the National Environmental Protection Agency (ISPRA), the National Association of Italian Municipalities (ANCI), and from waste hauling companies.<sup>18</sup> I collected a cross-section of price data directly from municipalities and companies upon request, and indirectly from municipal balance sheets. I created a database containing average or median prices over the analyzed time period depending on the acquired data. Detailed information on data management, variable denomination and descriptive statistics is presented in the Appendix B.

### 1.3 Summary Statistics

Table 1 compares key attributes of never-treated (non-PAYT) and treated (PAYT) units before policy. The main outcomes are kilograms of per capita unsorted (UW), recycling waste (RW), and total (TW) which is the sum of the previous two. Full summary statistics of the included attributes are presented in Tables 8 and 9 in Appendix B. Relevant predictors are discussed in the next section.

Table 1: Summary statistics for key attributes of never-treated vs. treated municipalities before policy. Waste amounts are measured in kilograms (kg) per capita (p.c.).

|             |                              | <u>Never-treated</u> |       | <u>Treated (before)</u> |       |
|-------------|------------------------------|----------------------|-------|-------------------------|-------|
| Obs. 19,448 |                              | Mean                 | Sd    | Mean                    | Sd    |
| Outcomes:   | Recycling Waste (RW) kg p.c. | 233.9                | 82.84 | 296.8                   | 89.9  |
|             | Unsorted Waste (UW) kg p.c.  | 223.5                | 120.5 | 216.0                   | 115.1 |
|             | Total Waste (TW=RW+UW)       | 457.4                | 127.0 | 512.7                   | 142.9 |
| Costs €:    | RW unit costs per kg         | 0.18                 | 0.12  | 0.16                    | 0.08  |
|             | UW unit costs per kg         | 0.28                 | 0.19  | 0.29                    | 0.16  |
|             | RW management costs p.c.     | 40.34                | 22.48 | 47.73                   | 24.31 |
|             | UW management costs p.c.     | 53.83                | 34.26 | 50.01                   | 23.88 |
| Covariates: | Income p.c. (x €1,000)       | 13.9                 | 2.3   | 14.2                    | 2.0   |
|             | Pop. share with college deg. | 0.09                 | 0.03  | 0.09                    | 0.03  |
|             | Distance to PAYT city (km)   | 50.0                 | 63.2  | 33.8                    | 38.0  |
|             | (... up to 90 variables)     | (...)                | (...) | (...)                   | (...) |

<sup>18</sup>E.g., Aimag, Ascit, Clara, Dolomiti Energia, Hera, Iren, SEAB, Soraris, etc.

Households generate relatively more RW than UW on average. Municipal unit costs are largely heterogeneous, and managing one unit of UW costs on average twice as much as RW. Costs for UW are slightly higher also in per capita terms. Per capita costs indicate that one individual on average spends about €100 per year for waste services, and pays the most for unsorted waste.<sup>19</sup> This implies that UW reductions that do not translate into RW increases will potentially drive cost savings for both households and municipalities.

Comparing never-treated and treated units before adoption reveals that the latter recycle more on average. This is not surprising as adopting PAYT where households are used to recycle helps preventing policy adverse effects (Bueno & Valente, 2019). PAYT is implemented in municipalities with slightly higher income and comparable education levels. Units introducing PAYT, the treatment, in a certain year are much closer to units which already had the treatment in previous years, suggesting, e.g., information dissemination.

Focusing on treated municipalities, Table 2 compares per capita waste amounts and municipal waste management costs before and after policy.

Table 2: Before-and-after comparison of waste outcomes and management costs of treated units.

|            |                              | <u>Before</u> |       | <u>After</u> |       |
|------------|------------------------------|---------------|-------|--------------|-------|
| Obs. 1,164 |                              | Mean          | Sd    | Mean         | Sd    |
| Outcomes:  | Recycling Waste (RW) kg p.c. | 296.8         | 89.90 | 337.9        | 92.10 |
|            | Unsorted Waste (UW) kg p.c.  | 216.0         | 115.1 | 115.9        | 80.90 |
|            | Total Waste (TW=RW+UW)       | 512.7         | 142.9 | 453.8        | 127.2 |
| Costs €:   | RW management costs p.c.     | 47.73         | 24.31 | 54.29        | 20.20 |
|            | UW management costs p.c.     | 50.01         | 23.88 | 40.31        | 18.11 |
|            | RW unit costs per kg         | 0.16          | 0.08  | 0.17         | 0.07  |
|            | UW unit costs per kg         | 0.29          | 0.16  | 0.43         | 0.26  |

As expected, UW and TW decrease, and RW increases after PAYT. This leads to higher costs of RW and lower costs of UW per capita. Unit costs of UW increase after policy suggesting possible economies of scale. Yet, large standard deviations indicate that average effects mask important heterogeneities across municipalities.

<sup>19</sup>Municipal p.c. costs approximate household expenditures as, in Italy, municipalities are cost minimizers such that fee revenues finance total waste management costs, a principle known as budget balance.

Table 3 reports price summary statistics by policy year. Prices range between 1 cent and 18 euro cents per liter of UW, for a total of 45 price levels. The average price is 8 cents, and the standard deviation is 0.05. Performing Wilcoxon and Kolmogorov-Smirnov tests (1945; 1971) shows that price distributions do not statistically differ across policy years.

Table 3: Price summary statistics for PAYT municipalities by policy year.

| Treatment              | Year         | Mean  | Sd    | Min  | P50   | P75  | Max  |
|------------------------|--------------|-------|-------|------|-------|------|------|
| PAYT price € per liter | 1 (obs. 194) | 0.073 | 0.046 | 0.01 | 0.066 | 0.13 | 0.18 |
|                        | 2 (obs. 161) | 0.080 | 0.046 | 0.01 | 0.079 | 0.13 | 0.18 |
|                        | 3 (obs. 125) | 0.088 | 0.044 | 0.01 | 0.090 | 0.13 | 0.18 |

Using prices and UW post-policy, I calculate household variable costs in each municipality.<sup>20</sup> In the third policy year, households in high-price municipalities (>13 cents) pay on average €176 per capita. In low-price municipalities (<3 cents), they pay on average €21 per capita. The magnitude of this gap raises several questions. After partialling out confounding factors, do reactions in municipalities setting high versus low prices differ? What role play income effects in driving these differences? More broadly, how do behaviors toward recycling and waste avoidance adjust to a decrease in unsorted waste?

## 1.4 Predictors

Accounting for observable determinants of price adoption and waste generation requires adjustments for many sorts of covariates. I include waste generation and policy adoption determinants for a total of 90 municipal attributes which can be grouped into six categories: socio-economic, geographic, and political variables, neighborhood effects, pre-policy waste amounts and management costs.

*Socio-economic characteristics.*—This group of covariates proxies for heterogeneous household opportunity costs of time and space. Above all, the literature highlights income and education levels.<sup>21</sup> The latter is measured by

<sup>20</sup>I use waste haulers' data for liter-to-kg conversion. If missing, I use the median value.

<sup>21</sup>See, e.g., Callan & Thomas (2006); Kinnaman (2006); Jenkins *et al.* (2003); Miranda & Bauer (1996); Van Houtven & Morris (1999); Richardson & Havlicek (1978); Grossmann *et al.* (1974); Wertz (1976).

the population share with graduate degree or higher, and elementary degree or none. Income effects are indeterminate a priori (Callan & Thomas, 2006). Higher-income households may be less elastic due to, e.g., lower budget constraints and higher time opportunity costs. Both education and income may proxy for citizens' demands for environmental quality (Dunlap *et al.*, 2000). These variables, therefore, may drive both policy adoption and effects. Further, I control for demographics as, e.g., average house and household size, home ownership, age structure, and tourism intensity. Labor market attributes proxy for, e.g., time spent at home. I include the share of unemployed and out-of-the-labor force population, and indices of labor market activity, commuting intensity, and social deprivation.

*Geographic characteristics.*—I control for the distance of each municipality to waste incinerators, landfills, and hazardous waste treatment facilities. Proximity to waste sites may induce, for instance, lower waste generation through households' pollution awareness. In addition, distant waste sites may cause high transportation costs and, therefore, correlate with price adoption decisions. I further distinguished communities by urbanization levels, and regional or provincial seats. These variables proxy for differences in administrative capacity, recycling infrastructure, and PAYT system (Gradus *et al.*, 2019).

*Neighborhood effects.*—Vicinity to other PAYT municipalities likely matters for two reasons. First, information dissemination: municipalities may adopt PAYT to mimic successful neighbors' policies (Allers & Hoebe, 2010). Second, consortia effects: nearby municipalities often share the same waste hauling company to save, e.g., collection costs. As a proxy, I control for the distance to the closest municipality that implemented PAYT in prior years. Hence, this variable takes similar values for neighboring municipalities that adopt PAYT and belong to the same consortium.

*Political variables.*—Policymakers may adopt PAYT in response to citizens' demand for, e.g., better service quality, lower waste charges, and fairer waste pricing (Dijkgraaf & Gradus, 2009; Batllell & Hanf, 2008). Public engagement may, in turn, impact policy acceptance and reactions. To proxy for political participation, I include municipal level voter turnouts in the 2013 Italian



general election. Political polarization and lack of social cohesion may also impact policy adoption and effectiveness. Protest votes and extreme ideology are proxied by vote shares for big tent parties, and for extreme left- and right-wing parties. Mayor characteristics can matter as well. Newly elected, young mayors may be willing to invest in waste technology; citizens may elect green mayors putting PAYT in their agenda; locally born mayors of regional parties may rely on large consensus and push reforms. To control for this, I add mayor’s age, term length, party, and place of birth.

*Pre-policy waste generation.*—Lagged waste generation may drive policy adoption as well as effect heterogeneity (Kinnaman, 2006). In fact, lagged waste outcomes reflect initial opportunity costs of household recycling and waste avoidance, and account for existing differences in the recycling infrastructure, e.g., at the curb. The waste generation history also contains information about unobservables such as motivation and experience in waste reduction, and pro-environmental attitudes (Bueno & Valente, 2019). Including lagged waste amounts, therefore, accounts for possible time-varying effects of fixed unobservables.

*Pre-policy waste management costs.*—Price adoption decisions largely rely on municipal cost levels, both in per capita and per kg terms. If salient to the household, lagged per capita costs may also impact household waste generation. Cost variables include budgetary costs for waste collection (e.g., labor), disposal (e.g., machinery and land), transportation (e.g., trucks), treatment (e.g., store and transform), and administrative services, net of recycling revenues from selling products and energy recovery.

## 2 Theory and Empirical Framework

### 2.1 Theoretical Predictions

I am interested in how a price on household unsorted waste disposal affects waste generation and social welfare. I show that household responses on recycling and waste avoidance behaviors are ambiguous from a theoretical viewpoint, and depend on the household cost structure as well as on whether these

behaviors are substitutes or complements in household preferences. I provide simple theoretical predictions that account for effect heterogeneity in a possibly large number of household characteristics. Empirical analysis is, therefore, required to gauge the magnitude and the direction of price elasticities, as well as the impact of relevant socio-economic determinants. For the sake of simplicity, the theoretical model falsely assumes linear demands. In the empirical part, however, I will control for any potential nonlinear effect of prices. Finally, in line with the empirical setting, my discussion here will show that understanding policy effects on recycling and waste avoidance is crucial for welfare analysis.

*Heterogeneous demand responses to PAYT.*—By pricing unsorted waste generation, PAYT provides monetary incentives that decrease the relative price (opportunity cost) of two other waste disposal behaviors: waste avoidance and recycling. In this way, prices alter households’ optimization problem regarding waste generation. To formalize these arguments, I derive a utility maximizing model of household consumption that highlights heterogeneous opportunity costs of recycling and waste avoidance behaviors. Based on comparative statistics, my simple formulas show that reactions to prices crucially depend on household marginal costs of these behaviors, as well as behaviors’ substitutability or complementarity. All derivations are reported in Appendix A.

What makes recycling and waste avoidance more or less costly for households? The behavioral economic literature offers at least three (non mutually-exclusive) theories. Cecere *et al.* (2014) provides a first testable theory: recycling is more convenient than waste avoidance because the latter requires learning new techniques and relies on intrinsic motives. Taken to an extreme, households may find recycling so convenient that total waste increases, causing rebound effects (see, e.g., Hong & Adams, 1999, for evidence on Korea.)

Kinnaman (2006) provides a second testable theory: households recycling the most before policy would find it difficult to reduce waste substantially further after policy. Thus, waste habits pre-policy matter. Time opportunity costs and income levels can matter as well, since busy households may not take the time to engage in additional recycling.

Lastly, analyzing how behaviors affect each other, D’Amato *et al.* (2016) provide a third testable theory: recycling and avoidance behaviors may be negatively correlated due to multi-tasking effects. This refers to households

being induced to focus on one behavior and reallocate some attention away from the other (Holmstrom & Milgrom, 1991). Alternatively, behaviors could be complements if waste prices crowd in pro-environmental preferences by, e.g., raising awareness of the waste pollution problem.

*PAYT and social welfare.*— I define policy social welfare effects as changes in social costs of waste management caused by PAYT. These are measured as the sum of municipal (private) plus environmental (external) costs of unsorted and recycling waste caused by PAYT. How do PAYT-induced changes in household waste generation affect social welfare? This depends on the relative size of social benefits from unsorted waste reductions versus social costs of recycling increases.

The welfare analysis presents at least three challenges. First, municipal level estimates of the environmental impacts of unsorted and recycling waste are typically not available. Second, available estimates are generally not monetized, and the monetizing literature is scarce. Third, while it seems reasonable to assume that PAYT has no impact on external costs, private costs may instead be affected through, e.g., (dis)economies of scale. I overcome these issues by relying on prior estimates of environmental impacts, and assessing PAYT effects on municipal unit costs of waste management directly from the data.

Unsorted waste disposal causes external costs associated to landfill and incineration.<sup>22</sup> A review of the life-cycle literature shows that external costs are constant and rather small. Based on prior estimates, Kinnaman *et al.* (2014) assume a baseline external marginal cost of \$15/ton for landfill disposal and \$30/ton for incineration. Relying on these estimates, I assume an average external costs of unsorted waste of €20/ton for unsorted waste.<sup>23</sup>

Recycling waste causes, to a large extent, external benefits.<sup>24</sup> These are due to saved pollution from the avoided extraction of virgin materials. Based on prior estimates (\$200/ton), I assume a marginal external benefit of €180/ton for recycling. This is further reduced to €120/ton to account for actual recov-

---

<sup>22</sup>External costs include pollution from climate change emissions and waste transportation as well as local disamenity externalities such as nuisance effects to neighboring properties (Kinnaman *et al.*, 2014).

<sup>23</sup>Data on amounts sent to landfills and incinerators are unavailable. Thus, I consider the average cost.

<sup>24</sup>The treatment process of recycling also causes external costs. Yet, benefits overcompensate costs.

ery rates of 66% in Italy (Ronchi, 2016).

In my sample, private unit costs of recycling are on average higher, though comparable to its external benefits. Differently, private unit costs of unsorted waste are on average twice as high as its external costs. This implies that unsorted waste reductions that do not translate into an increase in recycling are the driving source of social cost savings. In other words, household reactions towards waste avoidance matter for welfare.

## 2.2 Identification and Empirical Specification

As the previous section shows, a large number of variables come into play in the estimation of heterogeneous causal effects of PAYT policies. Therefore, I have to include a large number of municipal characteristics possibly explaining price adoption and effectiveness.

I follow the potential outcome approach (Rubin, 1974). Let  $(X_{it}, Y_{it}, P_{it})$  be the available data for municipality  $i = 1, \dots, n$  at time  $t = 2010, \dots, 2015$ , where  $X_{it} \in \mathbb{R}^d$  is a vector of  $d$  covariates,  $Y_{it}$  is the waste outcome for either unsorted (UW), recycling (RW), or total waste (TW), and  $P_{it} \in \mathcal{P} = [0; p_{max}]$  is the price (treatment) variable in year  $t$ . Henceforth, I omit the true subscript  $t$  for simplicity whenever possible. Note that prices range between zero (for all untreated years) and the maximum price set in treated years ( $p_{max}$ ). For every unit  $i$ , there is a set of potential waste outcomes  $Y_i(p)$ ,  $p \in \mathcal{P}$ , each being a random variable mapping a particular potential treatment,  $p$ , to a potential outcome such that  $Y_i = Y_i(p)$ . This is also referred to as the unit level dose-response function. For any municipality defined by a vector of characteristics  $X_i = x$ , I wish to estimate the individual treatment effect of unit  $i$ , which is defined as  $Y_i(p) - Y_i(0)$  and is, however, unobserved for any unit. Therefore, I will estimate the Conditional Average Treatment Effect (CATE) function  $\Delta(x) = \mathbb{E}[Y_i(p) - Y_i(0) | X_i = x]$  and the Conditional Average Price Effect (CAPE) function  $\delta(x) = \frac{\partial \mathbb{E}[Y_i(p) | X_i = x]}{\partial p}$  under the “canonical” assumptions of unconfoundedness and no spillovers, aka Stable Unit Treatment Value Assumption (SUTVA). I refer to Hirano and Imbens (2004) and Imai and van Dyk (2004) for a description of these assumptions. In essence, unconfoundedness requires to have enough controls - usually pre-treatment covariates and outcomes - so that, conditional on those controls, treatment assignment is as

good as randomized. In the case of multivalued treatment, this assumption writes  $Y_i(p) \perp P_i | X_i \ \forall p \in [0; p_{max}]$ , i.e., requires conditional independence to hold for each value of the treatment. Imbens (2000) referred to this as weak unconfoundedness, since it does not require joint independence of all potential outcomes.<sup>25</sup> While this assumption is not directly testable, I make it plausible by controlling for a large set of covariates and pre-policy outcomes. I further assess unconfoundedness in Section 3.4.1.

SUTVA excludes the possibility of interference between units and, given the observed covariates, allows to consider the potential outcomes of one unit to be independent of another unit's treatment status. To fulfill this assumption, I control for neighborhood effects, and I assess robustness to spillovers in untreated units (see Sections 1.4 and 3.4.1).

In order to fulfill unconfoundedness in the case of non-random price adoption, one needs to control for the sources of self-selection, i.e., capture the effect of  $X_i$  on  $P_i$ . Consider the partially linear waste outcome model:

$$Y_i = \mathbb{E}[Y_i(0)|X_i] + P_i\delta(X_i) + \epsilon_i(P_i). \quad (1)$$

Under unconfoundedness,  $\mathbb{E}[\epsilon_i(P_i)|X_i, P_i] = 0$ ; further,  $\mathbb{E}[Y_i(0)|X_i]$  represents the possibly non-linear direct effect of covariates on untreated outcomes. Note that  $X_i$  is of potentially very high dimension. Let  $\mathbb{E}[Y_i|X_i]$  and  $\mathbb{E}[P_i|X_i]$  be the conditional outcome and price mean, respectively. By means of algebraic transformations, model (1) can be rewritten as:

$$Y_i - \mathbb{E}[Y_i|X_i] = (P_i - \mathbb{E}[P_i|X_i])\delta(X_i) + \epsilon_i(P_i), \quad (2)$$

where  $\delta(X_i)$  identifies the CAPE as the effect of the leftover price variation  $P_i - \mathbb{E}[P_i|X_i]$  on the leftover outcome variation  $Y_i - \mathbb{E}[Y_i|X_i]$  not explained by the observed covariates. The standard partially linear model considers solely the case of constant treatment effects (Robinson, 1988). Yet, Nie & Wager (2019) study identification of model (2) for flexible heterogeneous treatment effect estimation via machine learning approaches. In particular, we can estimate  $\delta(x)$  in two steps: First, we separately estimate the nuisance components

---

<sup>25</sup>Yet, it is difficult to think of applications where the weaker form would be plausible but the stronger form would not be. Differences between the two are rather conceptual (see Imbens, 2000, for details).

$\mathbb{E}[Y_i|X_i = x]$  and  $\mathbb{E}[P_i|X_i = x]$ . Second, we plug in their fitted values to obtain  $\hat{\delta}(x)$  by regressing residualized outcomes on residualized prices. In this step, we do not consider all residuals as equally important, and we estimate a local version of model (2) that gives more weight to those residuals in the neighborhood of  $x$  (see the next subsection for details).

Residual-on-residual regression methods, also known as R(esidualized)-learning in high-dimensional settings, make the parameter estimate insensitive to small errors in the nuisance components, thus improving its robustness. Note that, as an alternative approach, one could directly estimate model (1) by including the high-dimensional covariate set. However, regularization of  $\mathbb{E}[Y_i(0)|X_i]$  would cause an especially large bias when covariates are correlated with prices (Athey *et al.*, 2017; Chernozhukov *et al.*, 2017).<sup>26</sup> Conversely, residualization removes the correlation of covariates with both prices and outcomes, rendering the estimator robust to the parametric form in which covariates are included. In particular, residualization makes the estimator “doubly robust”, i.e., as long as either the estimator for either propensity scores or conditional outcome expectation is consistent, the resulting estimator for the treatment effect is consistent (Athey & Imbens, 2017a).

*Estimation of the unit level causal effect of continuous treatment.*—I now discuss how to estimate the unit level predictions for the conditional expectations of interest: as mentioned above, I estimate first the generalized propensity score  $s(x) := \mathbb{E}[P_i|X_i = x]$  and the expected outcome marginalizing over treatment  $y(x) := \mathbb{E}[Y_i|X_i = x]$ , and finally the causal effect of continuous treatment  $\delta(x)$  or CAPE. I propose to use nonparametric machine learning methods that explicitly account for heterogeneity in the parameter estimation procedure. This allows to overcome three econometric issues. First, standard  $k$ -nearest neighbor/kernel matching methods are bound to fail, as the concept of neighbor vanishes in high-dimensions (Abadie & Imbens, 2016; Giraud, 2015): conversely, machine learning allows to handle the large covariate dimension. Second, nonparametric estimation allows to capture possibly complex interactions in a data-driven model specification which, therefore, allows to reduce the risk of model misspecification and ad-hoc model selection. Third, and most

---

<sup>26</sup>Without regularization, the inclusion of a large set of partly correlated predictors may lead to, e.g., variance inflation and incorrect signs.

importantly, machine learning allows to consistently estimate the full CAPE mapping; in turn, this allows to construct a policy targeting function mapping observed covariates to unit level causal effects. Among the class of adaptive (data-driven)  $k$ -nearest neighbor matching estimators, I implement the Random Forest (RF) method developed in Athey *et al.* (2019) that generalizes the original algorithm of Breiman (2001) by adapting to the problem of both prediction and heterogeneous treatment effect estimation.

Breiman’s RFs make predictions as an average of trees, as follows (see also, for a more comprehensive treatment Scornet & Biau, 2016): (1) For each tree  $b = 1, \dots, B$ , draw a random subsample of training data  $s_b^{tr} \subseteq \{1, \dots, n\}$ ; (2) Estimate a tree via recursive partitioning on each such subsample of the data:<sup>27</sup> the resulting tree is a sequence of binary regions partitioning the covariate space and grouping observations in the bottom regions called leaves  $L_b$ ; (3) Make predictions of  $y(x)$  in the leaf containing the set of units  $\{i : X_i \in L_b(x), i \in s_b^{tr}\}$ , and average predictions made by individual trees:

$$\hat{y}(x) = \frac{1}{B} \sum_{b=1}^B \sum_{n=1}^N \frac{Y_i \mathbb{1}(X_i \in L_b(x), i \in s_b^{tr})}{|i : X_i \in L_b(x), i \in s_b^{tr}|} \quad (3)$$

where  $L_b(x)$  denotes the leaf of the  $b^{th}$  tree containing the training sample  $x$ . In the case of out-of-sample prediction, I estimate  $y(x)$  using  $\hat{y}^{-i}(X_i)$ , based on considering only those trees  $b$  for which  $i \notin s_b^{tr}$ .<sup>28</sup>

Breiman’s RF is understood as an ensemble method averaging predictions made by individual trees. However, as shown in Athey *et al.* (2019), we can equivalently think of RF as an adaptive kernel method. For instance, we can re-write the forest prediction (3) as  $\hat{y}(x) = \sum_{i=1}^n w_i(x) Y_i$  where  $w_i(x)$  is a data-adaptive kernel that measures how often the  $i^{th}$  training unit falls into the same leaf as  $x$ . Weights sum up to 1, and define the forest-based adaptive neighborhood of  $x$  (see Athey *et al.*, 2019, for a formal definition). This kernel-based perspective of RFs allows to cast this method as an adaptive locally weighted

<sup>27</sup>The process is termed recursive because each subsample of the data can in turn be split an indefinite number of times until the splitting process terminates after a particular stopping criterion is reached.

<sup>28</sup>Without out-of-sample prediction, random forest estimators are asymptotically normal, since each estimate is derived by averaging estimates from many trees, but they overfit and do not converge at the square-root- $n$  rate, thus, are bias-dominated (Athey & Imbens, 2016; Mentch & Hooker, 2016).

estimator that first uses a forest to calculate a weighted set of neighbors for each test point  $x$ , and then solves a weighted version of the residual-on-residual regression (2). The resulting CAPE estimator writes:

$$\hat{\delta}(x) = \frac{\sum_{i=1}^n w_i(x)(Y_i - \hat{y}^{-i}(X_i))(P_i - \hat{s}^{-i}(X_i))}{\sum_{i=1}^n w_i(x)(P_i - \hat{s}^{-i}(X_i))^2}, \quad (4)$$

where  $\hat{s}^{-i}(X_i)$  are out-of-sample predictions of the propensity scores  $s(x)$ . Equation (4) defines the CAPE as a weighted residual-on-residual regression estimator with forest weights  $w(x)$  defining the nearest neighbors of  $x$  produced by different trees.<sup>29</sup> This distinguishes the CAPE from the APE estimator which is the unweighted version of (4).<sup>30</sup> Variance of  $\hat{\delta}(x)$  is estimated by evaluating the estimator on bootstrapped half-samples of the training data, also called bootstrap of little bags (Athey *et al.*, 2019; Sexton & Laake, 2009).

As for Breiman’s forests, estimation of a tree proceeds greedily, namely, the partitioning is done continuously by choosing the split-point and the covariate that attain the best fit (maximize or reduce some metric in that specific partition). In the case of generalized random forest, such criterion is chosen to maximize heterogeneity in the quantity of interest (outcomes, propensity scores, treatment effects) between partitions. For instance, the metric chosen for estimation of the CAPE function maximizes  $n_L n_R (\hat{\delta}_L - \hat{\delta}_R)^2$  where  $\hat{\delta}_{L,R}$  is the CAPE estimated in each binary (Left; Right) partition including a fraction of training units equal to  $n_{L,R}$ . Athey *et al.* (2019) discuss the design of computationally efficient splitting rule for generalized forests in more detail.

*Estimation of policy social welfare effects.*—For each municipality  $i = 1, \dots, n$  with attributes  $X_i = x$ , I define the policy Social Welfare Effect ( $SW E^{P>0}(x)$ ) in per capita euros as the sum of municipal (private) plus environmental (external) costs of unsorted and recycling waste caused by PAYT ( $P > 0$ ). The

---

<sup>29</sup>RFs have tuning parameters such as minimum leaf-size and penalties for imbalanced partitions. These are obtained via cross-validation as in Tibshirani *et al.* (2018), i.e., choosing the ones that make out-of-sample estimates of the regression errors minimized in the R-learning objective as small as possible.

<sup>30</sup>Motivated by the R-learning equation (2), I estimate the APE as  $\frac{\sum_{i=1}^n (Y_i - \hat{y}^{-i}(X_i))(P_i - \hat{s}^{-i}(X_i))}{\sum_{i=1}^n (P_i - \hat{s}^{-i}(X_i))^2}$ .



$SW E^{P>0}(x)$  writes:

$$- \left[ CATE_{UW}(x)(PC_{UW}^{P>0}(x) + EC_{UW}) + CATE_{RW}(x)(PC_{RW}^{P>0}(x) + EC_{RW}) \right] \quad (5)$$

$CATE_{UW}(x)$  and  $CATE_{RW}(x)$  are municipal level causal effects of waste prices on unsorted and recycling waste, respectively (in kg per capita).  $PC_{UW}^{P>0}(x)$  and  $PC_{RW}^{P>0}(x)$  are municipal level unit costs of, respectively, unsorted and recycling waste management under PAYT (in €/kg).  $EC_{UW}$  and  $EC_{RW}$  are the external marginal costs of unsorted waste and recycling, respectively, where the first is positive and the second is negative (in €/kg). As a result, a negative  $CATE_{UW}(x)$  (UW reductions) brings both private and external benefits, while a positive  $CATE_{RW}(x)$  (RW increases) brings private costs and external benefits. The net welfare effect is ex-ante unclear, and remains an empirical question.

While  $EC$  are assumed to be constant across units and rely on prior estimates,  $CATE(x)$  and  $PC(x)$  are estimated for all municipalities in the sample, treated and untreated.  $CATE(x)$  is estimated for each waste type (here, UW and RW) as  $p\hat{\delta}(x)$  with  $\hat{\delta}(x)$  being *statistically significant* CAPE estimates for the respective waste type.<sup>31</sup> For untreated units, the price  $p$  is assigned based on closest generalized propensity score prediction.

$PC^{P>0}(x)$  are observed for treated units (under treatment) but unobserved for untreated units. In order to predict counterfactual values for untreated units, I use the R-learning RF method described in Section 2.2. I regress all the covariates - including municipal attributes, lagged waste and cost values - and a policy indicator on municipal unit costs, running separate forests for UW and RW. I obtain a kernel function that maps municipal observables to policy causal effects on unit costs. I next use this function to predict unit costs for untreated units under PAYT. For each municipality defined by  $X_i = x$ , I assess if PAYT affects unit costs significantly. If this is the case, I replace  $PC^{P>0}(x)$  for untreated units with  $\hat{PC}^{P>0}(x)$ , their predicted counterpart under PAYT.

---

<sup>31</sup>As CAPE is the first derivative of the CATE, then an estimate of the CATE is just the integral of the estimated CAPE, i.e. of  $\hat{\delta}(x)$  (Torricelli-Barrow theorem).

### 3 Main results

I have explored from 500 to 10,000 trees in the RF, and treatment effect estimates become stable after 1,000 trees, thus, results are obtained using this value. Further, each tree is built with data for the same year to account for common shocks to all units. All trees are grown with cross-validated values for the number of randomly subsampled covariates, minimum leaf size, and penalty for imbalanced splits, namely, splits in which the size of parent and child node are very different are penalized. Additionally, since the treatment group is substantially smaller than the control group, each node is required to include a minimum number of both treated and control units, i.e., enough information about both factual and counterfactual to estimate the treatment effect reliably. For this reason, a penalization is imposed also to nodes including an unbalanced number of treated and control units. Following Athey *et al.* (2019), values for such parameters are obtained via cross-validation.<sup>32</sup> In order to credibly estimate policy causal effects, the common support condition for treated and untreated units needs to hold. Figure 10 in Appendix C provides evidence of such support.

#### 3.1 Behavioral responses

I begin my empirical analysis by estimating unit level causal effects of prices on demands for unsorted, recycling, and total waste per capita (UW, RW, TW, respectively). In this section, I focus the attention on elasticities in the third policy year in order to have a closer estimate to the long-term effects of PAYT.<sup>33</sup> I run separate random forests for each waste outcome: UW, RW, and TW (for robustness). Figure 2 presents the distribution of the estimated semi-elasticities of waste demands for all municipalities in the sample. All estimates are statistically different from zero (p-values  $< 0.01$ ). The hypothesis of effect homogeneity is rejected for all outcomes and policy years.<sup>34</sup>

---

<sup>32</sup>I use the software R-3.4.2 and the grf package version 0.10.0 (Tibshirani *et al.*, 2018).

<sup>33</sup>I analyze the dynamics of price effects in Section 3.2. One caveat is that these elasticities cannot be estimated for late adopters. Reassuringly, there are no statistically significant differences in the distribution of price effects between early and late adopters in the first two policy years.

<sup>34</sup>See Levene's tests (1960) and heuristics (Athey *et al.*, 2017) in Appendix D.2 and Table 11 therein.

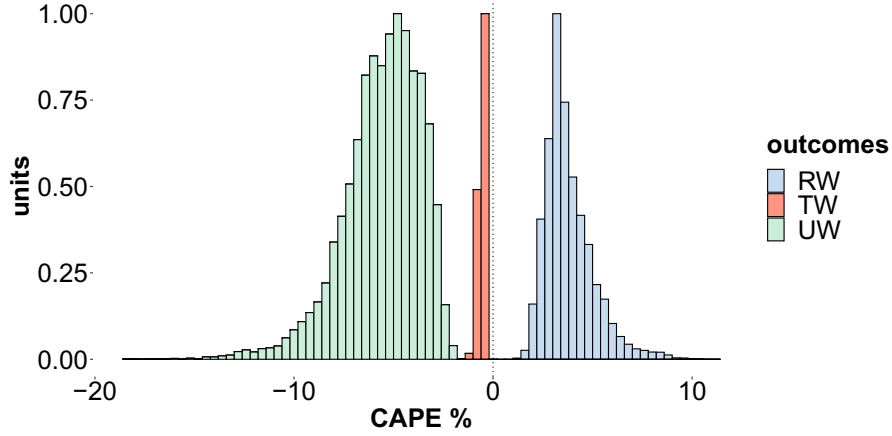


Figure 2: Unit level estimates of price semi-elasticities (CAPE) of waste demands. CAPE are measured as % changes of waste amounts for a one cent price increase.

The estimated price effects on waste demands for UW and RW are large in magnitude. Using the estimates to compute factual and counterfactual waste amounts implies that prices cause an annual UW decline of 50% on average. This reduction is mainly driven by an average RW increase of 32% and, to a lesser extent, by an average TW reduction of 5%.<sup>35</sup> The large substitution effect of UW with RW caused by the policy, therefore, represents a strong change in household waste behavior.

I proceed with the analysis of the sources of effect heterogeneity. In order to summarize the CAPE function, I linearly project the CAPE estimates onto price levels, a set of features, and their interactions (Tibshirani *et al.*, 2019). The literature on waste prices is especially interested in learning the elasticity of demand as a function of a few variables such as income or education. Thereby, the set of features includes relevant regressors capturing household opportunity costs of waste disposal: income, education, and pre-policy waste levels. Figure 3 plots fitted semi-elasticities (CAPE) across prices, *ceteris paribus*.<sup>36</sup> Pointwise estimates are obtained by fitting a smooth regression line. Shaded regions are the 95% confidence intervals.

<sup>35</sup>This corresponds to -110 kg (UW), +80 kg (RW) and -25 kg (TW) per capita. Thereby, estimates from separate forests are consistent with each other ( $CATE_{TW} \approx CATE_{UW} + CATE_{RW}$ ).

<sup>36</sup>Data inspection suggests quadratic demand curves. Thus, I fit a polynomial regression of order 2. Results for CAPE in kg are presented in Figure 11 in Appendix D.3.

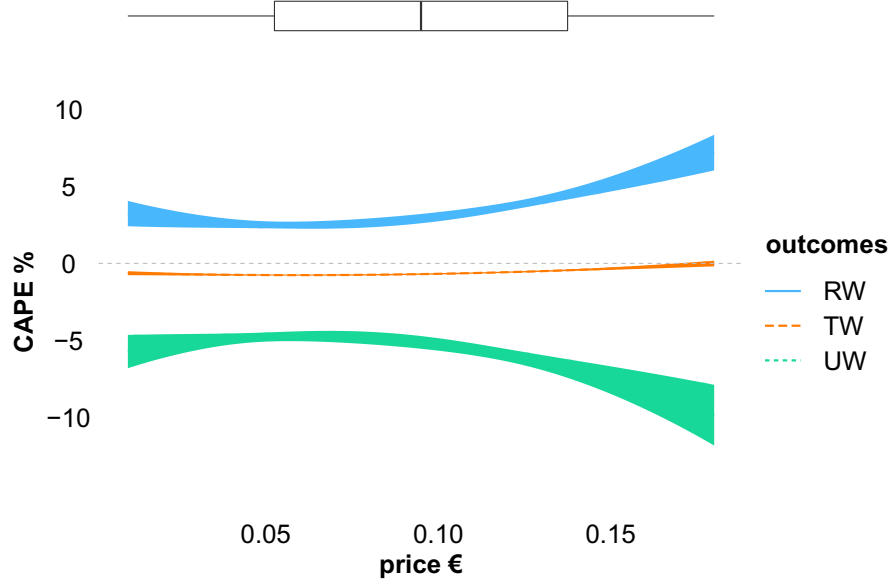


Figure 3: Fitted price semi-elasticities (CAPE) of waste demands at each price level.

Slopes of waste demands in high and low price regions differ. I use the median value of 9 cents to define these regions, indicated by the vertical line in the boxplot above the image. While at high prices elasticities are increasing, elasticities are rather constant at low prices. The estimated elasticity variation at low prices is 4% of the variation at high prices. All other things equal, a one cent price increase reduces unsorted waste by 5 to 10% (vs. 4.7-5.7% at low prices), increases recycling by 2 to 6% (2.5-3.2%) and reduces total waste by 0.1 to 0.7% (0.6-0.8%). Further, household behavior at high prices sheds new light on the mechanism behind UW reductions: increasing UW reductions reveal a negative correlation between RW and TW reactions, suggesting that recycling and waste avoidance are substitutes in individual preferences.<sup>37</sup> Thus, higher prices do not further reduce total waste, instead they simply reallocate waste to the recycling pile.

Having established effect heterogeneity across prices, I analyze possible income effects. Figure 4 plots fitted semi-elasticities by income level, *ceteris paribus*.

<sup>37</sup>This result confirms case (i) among the theoretical predictions reported in Appendix A. Estimates of the elasticity of substitution between recycling and avoidance are provided in Section 3.3.

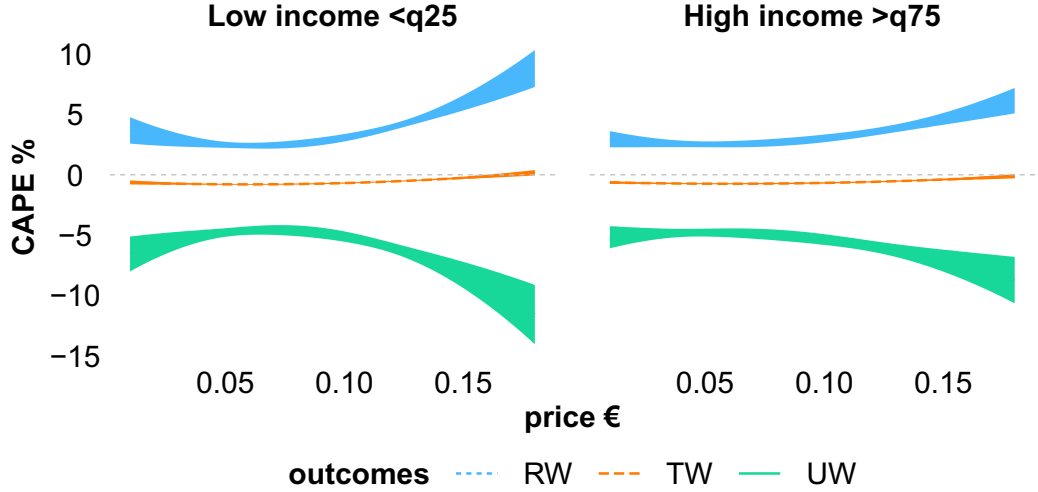


Figure 4: Fitted price semi-elasticities (CAPE) of waste demands by income levels. Thresholds q25 and q75 indicate first (€13k) and third (€16k) quartiles of annual income per capita.

Low-income municipalities are on average one percent point more elastic than high-income municipalities.<sup>38</sup> Thus, high prices enter budget constraints. However, income effects are not statistically significant at low prices. This means that high- and low-income households are similarly elastic, and possible differences in time opportunity costs do not trigger heterogeneous reactions.

Findings can be interpreted as follows: on the one side, no income effects and constant elasticities at low prices are consistent with nudge theories where small incentives have substantial behavioral impacts without budgetary incidence (Farhi & Gabaix, 2020). This interpretation is supported by the low monetary benefits of reducing UW at low prices: households reduce UW by 13.6% in the first price quartile (3 cents) and save less than one euro. On the other side, income effects at high prices provide evidence against possible regressive effects of waste prices. This result is new to the literature as previous studies typically consider the effects of low prices.<sup>39</sup>

Looking at heterogeneity across education levels, I estimate no significant

<sup>38</sup>In particular, at high prices, semi-elasticities for low-income municipalities are on average -7.5% for UW, 5.2% for RW, and -0.5% for TW, *ceteris paribus*.

<sup>39</sup>The highest price in, e.g., Fullerton & Kinnaman (2000) is \$2.18 per 32-gallon bag (120L) or \$0.02/L. This data include prices up to €0.18/L. See Bel & Gradus (2016) for a review of previous studies.

effects, suggesting that education plays, if any, a relatively minor role. Conversely, recycling habits before policy matter, as predicted by Kinnaman (2006): households recycling little before policy are more elastic.<sup>40</sup>

The literature typically estimates average point-elasticities smaller than one. For comparison, I compute point-elasticities at mean values and find similar values (see Table 10 in Appendix D.1). However, converting slope estimates into point-elasticity estimates may not be useful since price elasticity varies along the demand curve.

### 3.2 Welfare analysis

Welfare effect heterogeneity is driven by both demand and cost effects. In order to understand why municipalities may refrain from policy adoption, this section analyzes welfare effects in the short- versus longer-run. Demand effects (CATE) on UW are lowest in the first policy year. This is consistent with economic theory showing that elasticities are typically lower in the short-run because fewer alternatives are available (Usui, 2009). UW reductions are largely driven by increased recycling and, to a smaller extent, by waste avoidance. Further, behavior changes toward waste avoidance are relatively smaller and slower. This is consistent with behavioral economic predictions by e.g. Cecere *et al.* (2014) who argues that waste avoidance requires more effort and learning than recycling.<sup>41</sup> Figure 5 plots the distribution of statistically significant demand effects on TW by policy year.

---

<sup>40</sup>See Figure 12 in the Appendix D.3.

<sup>41</sup>For instance, households need to learn how to reuse and buy products with less packaging.

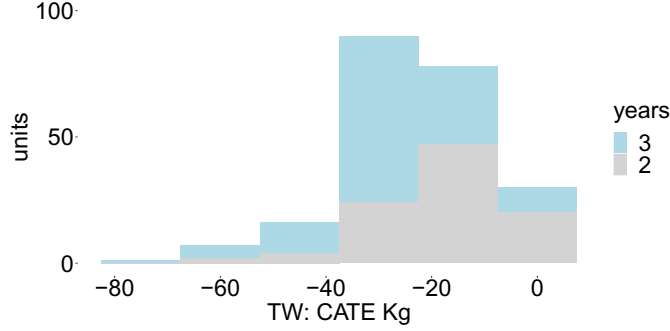


Figure 5: Unit level estimates of causal effects (CATE) on total waste demands in the second and third policy year (p-values  $< 0.05$ ). The y-axis refers to the number of treated municipalities.

In year 1, there are no statistically significant effects on waste avoidance (total waste reduction). In year 2 (grey), waste avoidance grows in a fraction of municipalities, and increases significantly for all municipalities in year 3 (blue).

Next, I flexibly estimate policy causal effects on unit management costs for UW and RW via random forests: overall, I find no economies of scale, confirming results on Italian municipalities by e.g. Abrate *et al.* (2014). In a fraction of municipalities, estimates suggest economies of scale for UW and RW because their unit cost increases (decreases) when UW (RW) reduces (increases).

Finally, I compute social welfare effects by combining CATE with management and pollution cost estimates. Table 4 shows social welfare effects (€ p.c.) by policy year.

Table 4: Summary statistics for Social Welfare Effects (SWE). Estimated private (municipal), external (environmental), and social cost effects per capita (p.c.) by policy year.

| Year | SWE € p.c. | Mean  | Sd    |
|------|------------|-------|-------|
| 1    | Private    | 0.13  | 10.16 |
|      | External   | 6.87  | 5.33  |
|      | Social     | 7.01  | 10.35 |
| 2    | Private    | 3.14  | 7.10  |
|      | External   | 5.96  | 4.75  |
|      | Social     | 9.11  | 9.08  |
| 3    | Private    | 20.17 | 16.28 |
|      | External   | 4.45  | 3.52  |
|      | Social     | 24.62 | 19.71 |

Especially in year 1 and 2, welfare effects can be negative. As feared, higher recycling can indeed increase costs for society. In year 3, however, PAYT generates welfare benefits in most municipalities. Average benefits are €25 per capita, which is about one fourth of what municipalities spend for waste management per person on average. These benefits are mainly due to private costs savings. As management unit costs are largely unaffected, savings come from UW reductions that do not translate into higher RW. In other words, welfare benefits are largely driven by waste avoidance.

As higher prices promote increased recycling rather than waste avoidance, the largest welfare benefits occur in municipalities setting low prices after three years of adoption. Figure 6 shows social welfare effects at different quartiles of the price distribution in the third policy year, when waste avoidance is highest.

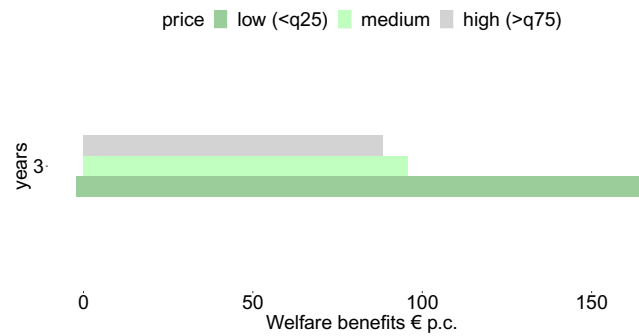


Figure 6: Social welfare effects (largely, welfare benefits) at different price levels after three years of adoption. First and third quartiles correspond to 3 and 13 cents, respectively.

This implies that (i) higher prices may be less desirable from a social cost perspective, although they cause larger unsorted waste reductions, and (ii) targeting waste avoidance in high-price municipalities could improve welfare benefits substantially.

The final implication of this paper may help local governments to overcome their initial fears regarding the implementation of PAYT policies. This paper provides evidence that short-term costs may indeed increase due to higher recycling,<sup>42</sup> but these costs are outrun by the long-term savings due to reductions in overall waste. This new information may help municipalities to overcome

<sup>42</sup>Especially if municipalities ignore global environmental benefits of recycling, and consider only avoided local pollution due to unsorted waste reductions.



their hesitation in adopting waste charges, and to increase policy benefits by focusing on waste avoidance targets.

### 3.3 Welfare simulations

Using random forests, I estimate a kernel function that maps municipal observables to policy causal effects. I next use this function to estimate overall causal effects on waste demands and social welfare if all municipalities in the sample were to implement PAYT.

Figure 7 plots the predicted price semi-elasticities of RW and TW demands after three years of adoption. As expected, color differences in the same area indicate that higher recycling is associated to lower avoidance and vice versa.

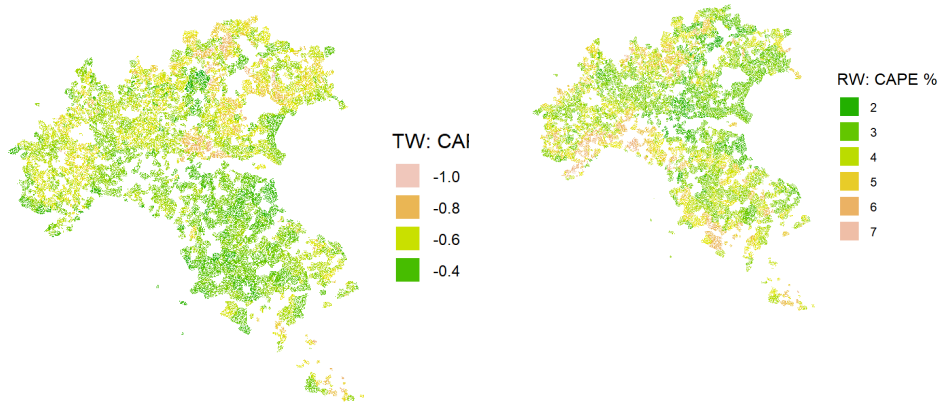


Figure 7: Predicted semi-elasticities (CAPE) on total waste and recycling demands for all municipalities three years after policy. Higher elasticities on waste avoidance and recycling indicated by warmer colors.

Figure 8 shows the negative correlation between price elasticities of recycling and waste avoidance aggregated by percentiles.

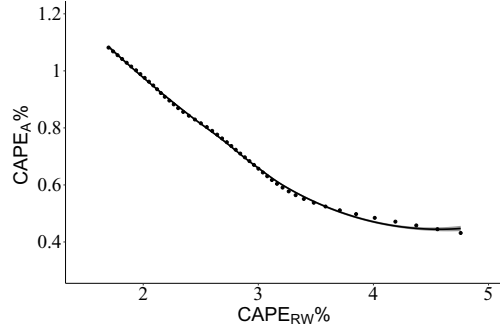


Figure 8: Semi-elasticities (CAPE) on recycling ( $x$ -axis) plotted against semi-elasticities on avoidance ( $y$ -axis). Note that  $-\text{CAPE}_{TW} = \text{CAPE}_A$ . A smooth line is fitted through percentile values.

The elasticity of substitution between behaviors is not constant and does not lead to rebound effects on total waste (i.e., a negative  $\text{CAPE}_A$ ). Using local linear regression (Li & Racine, 2004), I estimate that one percent point increase in the price semi-elasticity of recycling reduces that of waste avoidance between 0.03 and 0.4 percent points. This result is consistent with theories predicting multi-tasking effects between recycling and waste avoidance behaviors (e.g. D’Amato *et al.*, 2016), and provides new evidence on the magnitude of these effects.

Despite substitutabilities between recycling and avoidance behaviors, welfare simulations show that most municipalities would benefit from PAYT adoption. Table 5 shows predicted welfare effects by policy year if all municipalities were to implement PAYT.

Table 5: Summary statistics for predicted Social Welfare Effects (SWE). Estimated private (management), external (environmental), and social cost effects per capita (p.c.) by policy year.

| Year | SWE € p.c. | Mean  | Sd    |
|------|------------|-------|-------|
| 1    | Private    | 1.45  | 10.90 |
|      | External   | 7.05  | 4.64  |
|      | Social     | 8.50  | 10.51 |
| 2    | Private    | 3.95  | 9.33  |
|      | External   | 6.90  | 4.90  |
|      | Social     | 10.86 | 10.54 |
| 3    | Private    | 23.85 | 18.33 |
|      | External   | 5.44  | 4.20  |
|      | Social     | 29.29 | 21.97 |

Compared to counterfactual municipal and pollution costs, UW reductions lead to benefits for both waste management (-24%) and the environment (-2.3%) on average over the policy years. Further, increased RW causes higher costs for waste managements (+21%) but also pollution reductions for the environment (-15%). In year 3, average welfare benefits per capita amount to €30, however, there is large variation. Figure 9 maps this variation: despite effect heterogeneity, waste prices cause welfare benefits for most municipalities.

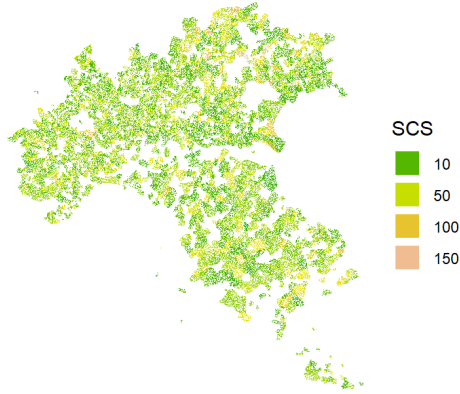


Figure 9: Distribution of predicted welfare benefits in the third policy year in € per capita.

### 3.4 Robustness and sensitivity

#### 3.4.1 Assessing assumptions: Unconfoundedness and SUTVA

*Anticipation effects.*—By conditioning on a large number of observables, this analysis aims to make the impact of leftover unobserved factors negligible and, thereby, unconfoundedness a plausible assumption. Although being per se untestable, unconfoundedness can be assessed by testing for significant effects on outcomes pre-policy (Athey & Imbens, 2017b). I run a RF including a binary treatment indicator, all covariates and available lags (up to order 2). Table 6 shows average price effects (APE) for three years pre-policy estimated as average treatment effects at the mean price level ( $ATE/\bar{P}$ ).

Table 6: Placebo APE at the mean price ( $ATE/\bar{P}$ ) from log-level (%) and level-level (kg) residual-on-residual regression (\*\*p-value < 0.01). Standard errors are clustered by year.

| Year | UW %  | kg    | s.e.  | RW %     | kg      | s.e.  | TW %    | kg      | s.e.  |
|------|-------|-------|-------|----------|---------|-------|---------|---------|-------|
| -3   | -5.63 | -2.90 | 9.81  | 6.07     | 12.5    | 10.54 | -2.34   | -11.3   | 21.33 |
| -2   | -6.07 | -16.8 | 17.98 | 5.16     | 12.6    | 8.42  | 2.68    | 15.5    | 22.27 |
| -1   | -8.29 | -16.9 | 13.26 | 21.59*** | 46.4*** | 9.53  | 9.60*** | 28.8*** | 10.27 |

Results suggest that unconfoundedness holds. Though, as expected, there is evidence of anticipation effects in the first pre-policy year. Before the official start, municipalities typically implement pilot programs where the new system is in place but households still pay according to the old regime of flat fees. This serves to promote recycling habits and minimize the risk of policy adverse effects. Estimates show a significant increase in RW that leads to higher TW, indicating that households largely substitute towards more recyclable products as, e.g., bulky items (Hong & Adams, 1999; Podolsky & Nestor, 1998). In terms of magnitude, the APE on TW approximately equals the sum of APE on RW and UW, however, a large standard error makes the latter statistically insignificant.

*System effects.*—I assess another possible threat to identification: unaccounted-for heterogeneity in system adoption.<sup>43</sup> In the sample, about 7% of treated units pay per weight, 16% per bag, 72% per emptying, and 5% have a mixed system. I first test for whether the included covariates can explain system adoption. Second, I test for significant differences in price elasticities between systems. Using all covariates, system choices are predicted by RF for classification. Reassuringly, observables can predict the adoption of volume systems correctly, out-of-sample, and misclassify 33% of weight systems only. Next, I test for effect heterogeneity. Results of pairwise mean comparison tests (Games & Howell, 1976) show that price elasticities do not statistically differ between systems.<sup>44</sup>

*Waste tourism.*—Violation of the no spillover assumption (SUTVA) could possibly come from waste tourism, namely, households may discharge their

<sup>43</sup>I cannot include system dummies in RF because, as for any policy indicator, I would force treated and untreated units into different neighborhoods.

<sup>44</sup>Results are reported in Table 13 in Appendix D.4.2.

waste for free in surrounding municipalities without PAYT. I geocode neighbors as never-treated units sharing a border with treated units. The remaining untreated units represent the control group. I find that PAYT has no significant effect on the waste amounts collected in surrounding municipalities on average. As a further robustness check, I re-estimate causal effects for PAYT municipalities excluding neighbors from the control group, finding no significant change in estimates. As in other empirical studies (e.g. Carattini *et al.*, 2018; Allers & Hoeben, 2010; Dijkgraaf & Gradus, 2004), waste tourism is not relevant on any significant scale.

### 3.4.2 Supporting analyses: Robustness to alternative assumptions

I compare my average estimates to *(i)* the binary treatment case, which assumes no effect heterogeneity in prices; *(ii)* the dynamic (event-study-like) difference-in-differences design, which assumes no effect heterogeneity, and selection based on time-constant variables; and *(iii)* penalized linear regression methods known as LASSO, which assumes no effect heterogeneity, sparsity and linearity in a high-dimensional set of covariates.

*Binary treatment case.*—Results show that mean elasticities (APE) on UW are overestimated by 28% on average (see Table 14 in Appendix D.4.3). Why does this happen? The forest kernel assigns similarity weights to units based on a residual-on-residual regression that ignores the price variation and uses a binary indicator instead. Hence, the algorithm may place units adopting high and low prices in the same neighborhood, and assign the same prediction to both units. However, I have shown that price responses at high and low prices are not the same, and the sources of effect heterogeneity differ.

*Difference-in-Differences.*—Results from a two-way fixed effect regression show that APE on UW are underestimated by 29% on average (see first panel of Figure 13 in Appendix D.4.3). The identifying parallel trend assumption is violated. There are at least three possible sources of bias. First, unaccounted-for time-varying effects on the outcomes (Bueno & Valente, 2019). Second, non-interpretable causal coefficients (Callaway & Sant’Anna, 2018). Due to the staggered policy adoption, APE represent weighted averages of effects computed using both treated and untreated outcomes, which may not allow for

a comparison with my estimates. Third, falsely assuming policy homogeneity (Abraham & Sun, 2018). Intuitively, pre-trends can arise solely from effect heterogeneity because the coefficient on a given lead or lag can be contaminated by effects from other periods.

*LASSO.*—Similarly to RF, doubly robust LASSO regression estimates the APE after removing the correlation of covariates with outcomes and prices. Yet, differently from RF, it assumes causal effect homogeneity and, typically, linearity in covariates (Athey *et al.*, 2017; Belloni *et al.*, 2017). I find that my APE estimates are almost insensitive to these assumptions when differences in covariates are effectively adjusted. Yet, this method does not allow to consistently estimate unit level causal effects and identify effect heterogeneity.

*Summary.*—Table 7 summarizes all APE estimates by estimation method. Residualized reg via RF is my main estimation method; residualized reg via RF (bin.) assumes no heterogeneity in prices; dynamic difference-in-differences (DiD) assumes constant APE and selection due to time-constant variables; double selection and residualized LASSO reg assume constant APE and linearity after effectively removing differences in covariates.

Table 7: Method comparison: APE estimates in the third policy year. Estimates represent average kg changes in waste for a one cent price increase.

| Method/APE estimates           | UW kg  | s.e. | RW kg | s.e. | TW kg | s.e. |
|--------------------------------|--------|------|-------|------|-------|------|
| residualized reg via RF        | -11.50 | 1.50 | 8.10  | 2.20 | -2.70 | 0.90 |
| residualized reg via RF (bin.) | -16.50 | 1.10 | 13.0  | 2.20 | -3.20 | 2.90 |
| dynamic DiD reg                | -7.40  | 0.30 | 2.70  | 0.70 | -4.70 | 0.50 |
| double selection LASSO reg     | -9.30  | 0.40 | 9.40  | 0.60 | -1.00 | 0.50 |
| residualized LASSO reg         | -9.00  | 1.30 | 8.80  | 1.00 | -1.00 | 1.30 |

## 4 Conclusions

This paper models and estimates causal effects of waste prices (PAYT) on recycling and avoidance behaviors by explicitly accounting for heterogeneity with machine learning methods. I analyze unique data with large variation in price levels and a high-dimensional set of municipal characteristics. I estimate municipal level price elasticities using a random forest and R-learning estimator

robust to confounding that affects outcomes and prices. By mapping municipal attributes to causal effects, I predict heterogeneous policy effects for all municipalities in the sample, treated or not. The welfare analysis is motivated by a theoretical model that predicts heterogeneous reactions on recycling and total waste. On the one hand, higher recycling brings environmental benefits. On the other hand, it increases municipal management costs. Theoretically, total waste reductions seem crucial to decrease municipal social costs substantially, and generate overall welfare benefits.

Disentangling effect heterogeneity in prices from other sources, I find that waste demands are nonlinear. Low prices, nudging households toward similar reactions, trigger large unsorted waste reductions, increase recycling, and decrease total waste. High prices make waste demands increasingly elastic, and induce recycling rather than waste avoidance. On average, policy causal effects amount up to -50% on unsorted waste, 32% on recycling, and -5% on total waste. Income effects and waste habits pre-policy well explain nonlinearities at high prices. I find that lower-income municipalities are more price elastic, which indicates that PAYT is not regressive. Despite large effect heterogeneity, the policy reduces social costs in most municipalities after three years of adoption, especially where total waste reductions are large. This implies that higher prices may be less desirable from a social cost perspective although they cause relatively larger unsorted waste reductions.

In sum, this study shows that price levels matter for household waste behavior as well as social welfare. Estimating heterogeneous responses on unsorted waste, and especially on recycling and total waste is crucial to assess overall welfare effects. Behavior changes towards waste avoidance, however, may not be immediate, and social costs may increase in the short-run. One upshot of this paper is that low prices can have a big impact on waste behavior and social welfare. Analyzing the long-term effects of PAYT policies presents an exciting research opportunity that I hope to tackle in future work.

## References

- Abadie, A., & Imbens, G. W. 2016. Large sample properties of matching estimators for average treatment effects. *Econometrica*, **74**(1), 235–267.
- Abbott, A., Nandeibam, S., & O’Shea, L. 2013. Recycling: Social norms and warm-glow

- revisited. *Ecological Economics*, **90**, 10–18.
- Abraham, Sarah, & Sun, Liyang. 2018 (Apr.). *Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects*. Papers 1804.05785. arXiv.org.
- Abrate, G., Erbetta, F., Fraquelli, G., & Vannoni, D. 2014. The Costs of Disposal and Recycling: An Application to Italian Municipal Solid Waste Services. *Regional Studies*, **48**(5), 896–909.
- Allers, M., & Hoeben, C. 2010. Effects of unit-based garbage pricing: A differences-in-differences approach. *Environmental Resource Economics*, **45**(3), 405–28.
- Athey, S., & Imbens, G. 2016. Recursive Partitioning for Heterogeneous Causal Effects. *PNAS*, **113**(27), 7353–7360.
- Athey, S., & Wager, S. 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, **113**(523), 1228–1242.
- Athey, S., & Wager, S. 2019. Estimating Treatment Effects with Causal Forests: An Application. *Observational Studies*.
- Athey, S., Imbens, G., Pham, T., & Wager, S. 2017. Estimating Average Treatment Effects: Supplementary Analyses and Remaining Challenges. *American Economic Review*, **107**(5).
- Athey, S., Tibshirani, J., & Wager, S. 2019. Generalized Random Forests. *The Annals of Statistics*, **47**(2), 1148–1178.
- Athey, Susan, & Imbens, Guido W. 2017a. Estimating Average Treatment Effects: Supplementary Analyses and Remaining Challenges. *American Economic Review: Papers and Proceedings*, **107**(5), 278–281.
- Athey, Susan, & Imbens, Guido W. 2017b. The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives*, **31**(2), 3–32.
- Batllell, M., & Hanf, K. 2008. The fairness of PAYT systems: Some guidelines for decision-makers. *Waste Management*, **28**, 2793–2800.
- Bel, G., & Gradus, R. 2016. Effects of unit-based pricing on the waste collection demand: A meta-regression analysis. *Resource and Energy Economics*, **44**, 169–182.
- Belloni, Alexandre, Chernozhukov, Victor, Fernández-Val, Ivan, & Hansen, Christian. 2017. Program evaluation and causal inference with high-dimensional data. *Econometrica*, **85**(1), 233–298.
- Bénabou, R., & Tirole, J. 2003. Intrinsic and Extrinsic Motivation. *The Review of Economic Studies*, **70**(3), 489–520.



- Bowles, S., & Polania-Reyes, S. 2012. Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature*, **50**(2), 368–425.
- Breiman, Leo. 2001. Random Forests. *Machine Learning*, **45**(1), 5–32.
- Brekke, K.A., Kipperberg, G., & Nyborg, K. 2010. Social Interaction in Responsibility Ascription: The Case of Household Recycling. *Land Economics*, **86**(4), 766–784.
- Buccioli, A., Montinari, N., & Piovesan, M. 2015. Do Not Trash the Incentive! Monetary Incentives and Waste Sorting. *The Scandinavian Journal of Economics*, **117**(4), 1204–1229.
- Bueno, M., & Valente, M. 2019. The Effects of Pricing Waste Generation: A Synthetic Control Approach. *Journal of Environmental Economics and Management*, **96**, 274–285.
- Callan, S., & Thomas, J. 2006. Analysing demand for disposal and recycling services: A system approach. *East Economic Journal*, **31**, 221–240.
- Callan, S. J., & Thomas, J. M. 1999. Adopting a unit pricing system for municipal solid waste: Policy and socioeconomic determinants. *Environmental Resource Economics*, **14**(4), 503–518.
- Callaway, B., & Sant’Anna, P.H.C. 2018. Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment. *DETU Working Papers* 1804.
- Carattini, S., Baranzini, A., & Lalive, R. 2018. Is Taxing Waste a Waste of Time? Evidence from a Supreme Court Decision. *Ecological Economics*, **148**, 131–151.
- Cecere, G., Mancinelli, S., & Mazzanti, M. 2014. Waste prevention and social preferences: the role of intrinsic and extrinsic motivations. *Ecological Economics*, **107**, 163–176.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robinson, J. 2017. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, **21**, C1–C68.
- Chi-ang, B., & Zheng, S. 2017. *Environmental Economics and Sustainability*. Chapter 6, Wiley Blackwell.
- Conover, W. J. 1971. *Practical Nonparametric Statistics*. New York: John Wiley & Sons.
- CONSEA. 2019. Report on waste pricing. *CONSEA Waste Services and Consulting*. <http://www.consea.eu/la-tariffa-puntuale-dieci-domande-e-risposte-per-capire-meglio-di-cosa-si-tratta/>.
- Croissant, Yves, & Millo, Giovanni. 2008. Panel data econometrics in R: The plm package. *Journal of Statistical Software*, **27**(2), 1–43.

- D'Amato, A., Mancinelli, S., & Zoli, M. 2016. Complementarity vs substitutability in waste management behaviors. *Ecological Economics*, 84–94.
- Davis, J.M.V., & Heller, S.B. 2017. Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs. *American Economic Review*, **107**(5), 546–50.
- Dijkgraaf, E., & Gradus, R. 2015. Efficiency effects of unit-based pricing systems and institutional choices of waste collection. *Environmental and Resource Economics*, **61**(4), 791–810.
- Dijkgraaf, E., & Gradus, R. H. J. M. 2004. Cost savings in unit-based pricing of household waste: The case of The Netherlands. *Resource and Energy Economics*, **26**(4), 353–371.
- Dijkgraaf, E., & Gradus, R. H. J. M. 2009. Environmental activism and dynamics of unit-based pricing systems. *Resource and Energy Economics*, **31**(1), 13–23.
- Driscoll, John C., & Kraay, Aart C. 1994. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, **80**(4), 549–560.
- Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. 2000. New Trends in Measuring Environmental Attitudes: Measuring Endorsement of the New Ecological Paradigm: A Revised NEP Scale. *Journal of Social Issues*, **56**(3), 425–442.
- Facchini, Marta. 2020. Rifiuti preziosi: tariffa puntuale e raccolta differenziata convengono. <https://altreconomia.it/tariffa-puntuale-raccolta-differenziata/>.
- Farhi, & Gabaix, X. 2020. Optimal Taxation with Behavioral Agents. *American Economic Review*, **110**(1), 298–336.
- Ferrara, I., & Missios, P. 2012. A Cross-Country Study of Household Waste Prevention and Recycling: Assessing the Effectiveness of Policy Instruments. *Land Economics*, **88**(4), 710–744.
- Freyaldenhoven, S., Hansen, C., & Shapiro, J. M. 2019. Pre-event Trends in the Panel Event-Study Design. *American Economic Review*, **109**(9), 3307–3338.
- Fullerton, D., & Kinnaman, C. T. 1996. Household responses to pricing garbage by the bag. *The American Economic Review*, **86**(4), 971–984.
- Fullerton, D., & Kinnaman, C. T. 2000. Garbage and recycling with endogenous local policy. *Journal of Urban Economics*, **48**(3), 419–442.
- Games, P. A., & Howell, J. F. 1976. Pair wise multiple comparison procedures with unequal n's and/or variances. *Journal of Educational Statistics*, **1**, 13–125.
- Gilli, M., Nicoli, F., & Farinelli, P. 2018a. Behavioural attitudes towards waste prevention and recycling. *Ecological Economics*, **154**, 294–305.

- Gilli, M., Mancinelli, S., & Nicolli, F. 2018b. *Household waste management: Some insights from behavioural economics*. Cham: Palgrave Pivot, Economics and finance collection.
- Giraud, Christophe. 2015. *Introduction to High-Dimensional Statistics*. Monographs on Statistics and Applied Probability 139, CRC Press - Taylor & Francis Group.
- Gobillon, Laurent, & Magnac, Thierry. 2016. Regional policy evaluation: Interactive fixed effects and synthetic controls. *The Review of Economics and Statistics*, **98**(3), 535–551.
- Gradus, R., Homsy, George C., Liao, Lu, & Warner, Mildred E. 2019. Which US municipalities adopt Pay-As-You-Throw and curbside recycling? *Resources, Conservation and Recycling*, **143**, 178–183.
- Grossmann, D., Hudson, J. F., & Marks, D.H. 1974. Waste generation models for solid waste collection. *Journal of the Environmental Engineering Division*, **100**(6), 1219–1230.
- Gulyas, A., & Pytka, K. 2019. *Understanding the Sources of Earnings Losses After Job Displacement: A Machine-Learning Approach*. CRC TR 224 Discussion Paper Series.
- Heckman, James J., Ichimura, Hidehiko, & Todd, Petra E. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, **64**(4), 605–654.
- Hirano, K., & Imbens, G.W. 2004. The Propensity Score with Continuous Treatments. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*.
- Holmstrom, B., & Milgrom, P. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law Economics and Organization*, **7**, 24–52.
- Hong, S., & Adams, R. M. 1999. Household responses to price incentives for recycling: Some further evidence. *Land Economics*, **75**(4), 505–14.
- Huang, J-C, Halstead, J. M., & Saunders, S. B. 2011. Managing municipal solid waste with unit-based pricing: Policy effects and responsiveness to pricing. *Land Economics*, **87**(4), 645–660.
- Imai, K., & van Dyk, D.A. 2004. Causal Inference with General Treatment Regimes: Generalizing the Propensity Score. *Journal of the American Statistical Association*, **99**, 854–866.
- Imbens, G.W. 2000. The role of the propensity score in estimating dose-response functions. *Biometrika*.
- Imbens, G.W., & Rubin, D.B. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences - Chapter 14 Assessing Overlap in Covariate Distributions*. Vol. III. Cambridge University Press.

- ISPRA. 2019. Urban Waste Report. *Italian National Institute for Environmental Protection and Research*.
- ISTAT. 2011. Population and Housing Census. *ISTAT - National Institute of Statistics*. Data available at <http://dati-censimentopopolazione.istat.it>.
- ISTAT. 2018. Data on Environmental Crimes. *ISTAT - National Institute of Statistics*. Data available at <http://www.istat.it/it/archivio/218648>.
- Jacobson, Louis S., LaLonde, R. J., & Sullivan, D. G. 1993. Earnings Losses of Displaced Workers. *The American Economic Review*, **83**(4), 685–709.
- Jenkins, R. R., Martinez, S. A., Palmer, K., & Podolsky, M. J. 2003. The determinants of household recycling: A material-specific analysis of recycling program features and unit pricing. *Journal of Environmental Economics and Management*, **45**(2), 294–318.
- Kahn, M.E. 2007. Do greens drive Hummers or hybrids? Environmental ideology as a determinant of consumer choice. *Journal of Environmental Economics and Management*, **54**(2), 129–145.
- Kinnaman, C. T. 2014. Understanding the Economics of Waste: Drivers, Policies, and External costs. *International Review of Environmental and Resource Economics*, **8**, 281–320.
- Kinnaman, T. C. 2006. Examining the justification for residential recycling. *Journal of Economic Perspectives*, **20**(4), 219–32.
- Kinnaman, T. C., T., Shinkuma, & Yamamoto, M. 2014. The socially optimal recycling rate: Evidence from Japan. *Journal of Environmental Economics and Management*, **68**(1), 54–70.
- Klueve, J., Schneider, H., Uhlendorff, A., & Zhao, Z. 2012. Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society*, **75**(2), 587–617.
- Kramer, C. Y. 1956. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, **12**(3), 307–310.
- Legambiente. 2017. Report on waste pricing policies. *Ecoforum Lombardia*. [http://www.legambiente.it/sites/default/files/docs/2\\_-\\_la\\_tariffa\\_puntuale\\_sistemi\\_e\\_metodi\\_di\\_applicazione\\_g.\\_ghiringhelli.pdf](http://www.legambiente.it/sites/default/files/docs/2_-_la_tariffa_puntuale_sistemi_e_metodi_di_applicazione_g._ghiringhelli.pdf).
- Levene, H. 1960. Robust Tests for Equality of Variances. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*.
- Li, Q., & Racine, J.S. 2004. Cross-Validated Local Linear Nonparametric Regression. *Statistica Sinica*, **14**(2), 485–512.

- Mentch, Lucas, & Hooker, Giles. 2016. Quantifying Uncertainty in Random Forests via Confidence Intervals and Hypothesis Tests. *Journal of Machine Learning Research*, **17**(26), 1–41.
- Miranda, M. L., & Bauer, S. 1996. The urban performance of unit pricing: An analysis of variable rates for residential garbage collection in urban areas. *The U.S. Environmental Protection Agency*. Report prepared for the Office of Policy, Planning, and Evaluation.
- Nie, X., & Wager, S. 2019. Quasi-oracle estimation of heterogeneous treatment effects. <http://arxiv.org/abs/1712.04912>.
- PartItalia. 2020. PAYT: The importance of the different systems (in Italian). Available at <http://www.raccoltapuntualefacile.com/tariffa-puntuale-limportanza-dei-diversi-modelli/>.
- Pigou, A. C. 1932. *Divergences between marginal social net product and marginal private net product*. Vol. II, Chapter IX. London: Macmillan and Co.
- Podolsky, M. J., & Nestor, D. V. 1998. Assessing incentive based environmental policies for reducing household waste disposal. *Contemporary Economic Policy*, **16**(4), 27–39.
- Richardson, R. A., & Havlicek, J. H. 1978. Economic analysis of the composition of household solid wastes. *Journal of Environmental Economics and Management*, **5**(1), 103–111.
- Rivers, N. 2017. Using nudges to reduce waste? the case of Toronto’s plastic bag levy. *Journal of Environmental Management*, **188**, 153–162.
- Robinson, P.M. 1988. Root-N-Consistent Semiparametric Regression. *Econometrica*, **56**(4), 931–954.
- Ronchi, Edo. 2016. Report on L’Italia del Riciclo. *FISE Unire and Fondazione per lo Sviluppo Sostenibile*. Available at <http://www.fondazionesvilupposostenibile.org>.
- Rubin, Donald B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **5**(5), 688–701.
- Scornet, E., & Biau, G. 2016. A random forest guided tour. *Test*, **25**(2), 197–227.
- Sexton, Joseph, & Laake, Petter. 2009. Standard errors for bagged and random forest estimators. *Computational Statistics and Data Analysis*, **53**(3), 801–811.
- Skumatz, L. 2008. Pay as You Throw in the U.S.: Implementation, Impacts, and Experience. *Waste Management*, **28**(12), 2778–2785.
- Smith, V. 1972. Dynamics of waste accumulation: Disposal versus Recycling. *Quarterly Journal of Economics*, **86**(4), 600–616.

- Thøgersen, J. 2006. Norms for environmentally responsible behaviour: An extended taxonomy. *Environmental Psychology*, **26**(4), 247–261.
- Tibshirani, J., Athey, S., Wager, S., Friedberg, R., Miner, L., & Wright, M. 2018. grf: Generalized Random Forests (Beta). R package version 0.10.1. Available at <http://CRAN.R-project.org/package=grf>.
- Tibshirani, J., Athey, S., & Wager, S. 2019. The GRF Algorithm. *Github repository*. Available at <http://grf-labs.github.io/grf/REFERENCE>.
- Usui, T. 2009. Does a rebound effect exist in solid waste management? *Critical Issues in Environmental Taxation*, **7**, 413–422.
- Van Houtven, G. L., & Morris, G. E. 1999. Household behavior under alternative pay-as-you-throw systems for solid waste disposal. *Land Economics*, **75**(4), 515–537.
- Wager, S. 2020. Designing Loss Functions for Causal Inference. *St. Gallen Workshop on Machine Learning*.
- Wertz, K. L. 1976. Economic factors influencing households’ production of refuse. *Journal of Environmental Economics and Management*, **2**(4), 263–272.
- Wilcoxon, Frank. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin*, **1**(6), 80–83.

## A Appendix: Theoretical predictions

*The model.*—Consider the household problem to maximize utility from consumption net of the disutility of avoidance and recycling efforts. The corresponding expression is:

$$\max_{w_A, w_R} : V(\cdot) = U(y - t(W - (w_A + w_R))) - C(w_A, w_R, W_0, W_R(W_0), X, Z) \quad (6)$$

where  $U(\cdot)$  is the utility from consumption,  $y - t(W - (w_A + w_R))$ , with  $W - (w_A + w_R)$  being the amount of unsorted waste subject to the PAYT price,  $t$ , which is taken as exogenous in the following derivations. Avoided ( $A$ ) and recycling ( $R$ ) waste amounts,  $w_i$  ( $i = A, R$ ), are generated from a potential waste amount,  $W$ , associated to the consumption of goods purchased with income  $y$ . Total waste is defined by the sum of unsorted and recycling waste, and equals  $W - w_A$ . Regarding consumption utility, the household is assumed to prefer more

consumption to less ( $\frac{\partial U(\cdot)}{\partial y - t(W - (w_A + w_R))} > 0$ ), waste avoidance and recycling to unsorted waste generation. This implies that marginal consumption utility with respect to waste avoidance and recycling is positive ( $\frac{\partial U(\cdot)}{\partial w_i} > 0$ ) and decreasing ( $\frac{\partial^2 U(\cdot)}{\partial w_i^2} < 0$ ).

Moreover, households get disutility,  $C(\cdot)$ , from time and effort spent in disposing  $w_i$ .  $C(\cdot)$  is a standard microeconomic cost function for which the following conditions hold:

$$\frac{\partial C(\cdot)}{\partial w_i} > 0; \quad \frac{\partial^2 C(\cdot)}{\partial w_i^2} > 0; \quad \frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R} \leq 0. \quad (7)$$

The first two conditions indicate that avoiding and recycling waste is costly, and marginal costs are increasing with the reduction of unsorted waste. The third relationship, importantly, captures the degree to which households are able to shift between disposal methods, pointing to either complementarities or substitutabilities between the two behaviors with  $\frac{\partial^2 C(\cdot)}{\partial w_A \partial w_C} < 0$  or  $\frac{\partial^2 C(\cdot)}{\partial w_A \partial w_C} > 0$ , respectively. Empirically, an indication of substitute (complement) behaviors may derive from the estimation of a negative (positive) correlation between individual price reactions on recycling and on avoidance.

The arguments of the cost function,  $C(\cdot)$  are current  $\{w_i\}$  and initial (lagged) waste  $\{W_0, W_R(W_0)\}$  with the latter accounting for waste reduction and recycling habits. Additionally,  $C(\cdot)$  depends on waste-reduction technology shifters ( $X$ ) which are exogenous socio-economic factors such as household income and education, and taste shifters ( $Z$ ) such as preferences for the environment. In particular, households may derive utility from waste avoidance and recycling due to the psychological reward of contributing to public good provision or “warm-glow”, and external rewards (peer approval) for pro-social and environmentally responsible actions.<sup>45</sup>

Solving problem (6) for  $w_i$  ( $i = A, R$ ) leads to the following first-order

---

<sup>45</sup>For a discussion see, e.g., Abbott *et al.* (2013); Bénabou & Tirole (2003); Brekke *et al.* (2010); D’Amato *et al.* (2016); Gilli *et al.* (2018a); Kahn (2007); Jenkins *et al.* (2003); Thøgersen (2006). The PAYT price may also enter  $C(\cdot)$  directly since it may partly crowd out households’ intrinsic and extrinsic motivation, increasing the disutility from waste recycling and avoidance (see, e.g., Bowles & Polania-Reyes, 2012; Cecere *et al.*, 2014; Chi-ang & Zheng, 2017; Ferrara & Missios, 2012, and citations therein). Yet, this section outlines a simpler model predicting PAYT heterogeneous effects.

conditions:

$$\frac{\partial V(\cdot)}{\partial w_i} = U'(\cdot)t - \frac{\partial C(\cdot)}{\partial w_i}, \quad (8)$$

with  $U'(\cdot)t$  being the marginal utility of avoiding or recycling waste, and  $\frac{\partial C(\cdot)}{\partial w_i}$  its marginal cost. Optimality conditions (8) for both  $w_A$  and  $w_R$  allow to derive the comparative statics about the impact of PAYT pricing ( $t$ ) on waste avoidance ( $w_A$ ) and recycling ( $w_R$ ), implying that:

$$\text{sgn}\left(\frac{\partial w_A}{\partial t}\right) = \text{sgn}\left[\frac{\partial^2 C(\cdot)}{\partial w_R^2} - \frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R}\right]; \text{sgn}\left(\frac{\partial w_R}{\partial t}\right) = \text{sgn}\left[\frac{\partial^2 C(\cdot)}{\partial w_A^2} - \frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R}\right] \quad (9)$$

where  $\text{sgn}(\cdot)$  is a function that extracts the sign of PAYT causal effects. Specifically, equations (9) mean that the sign of PAYT causal effects depends on complementarities/substitutabilities between waste reduction behaviors. As a result, theoretical predictions of PAYT causal effects are threefold:

- (i) PAYT causes avoidance and recycling to increase ( $\frac{\partial w_A}{\partial t} > 0$ ,  $\frac{\partial w_R}{\partial t} > 0$ ) if  $A, R$  are *complements* ( $\frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R} < 0$ ) or if  $A, R$  are *substitutes* ( $\frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R} > 0$ ) but substitutabilities are small such that  $\frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R} < \frac{\partial^2 C(\cdot)}{\partial w_A^2}$ .
- (ii) PAYT causes total waste and recycling to increase ( $\frac{\partial w_A}{\partial t} < 0$ ,  $\frac{\partial w_R}{\partial t} > 0$ ) if  $A, R$  are *substitutes* ( $\frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R} > 0$ ) and marginal costs of avoidance are steeper than marginal costs of recycling such that  $|\frac{\partial^2 C(\cdot)}{\partial w_A^2}| > |\frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R}| > |\frac{\partial^2 C(\cdot)}{\partial w_R^2}|$ .
- (iii) PAYT causes avoidance to increase and recycling to decrease ( $\frac{\partial w_A}{\partial t} > 0$ ,  $\frac{\partial w_R}{\partial t} < 0$ ) if  $A, R$  are *substitutes* and marginal costs of recycling are steeper than marginal costs of avoidance such that  $|\frac{\partial^2 C(\cdot)}{\partial w_R^2}| > |\frac{\partial^2 C(\cdot)}{\partial w_A \partial w_R}| > |\frac{\partial^2 C(\cdot)}{\partial w_A^2}|$ .

Formulas show that whether PAYT causes an increase in both pro-environmental behaviors depends on household marginal costs of recycling versus avoidance relative to the degree to which behaviors are substitutes or complements in individual preferences.

*Full derivations.*—Total derivatives of optimality conditions in (8) for  $i = A, R$  write:

$$\frac{\partial^2 V(\cdot)}{\partial w_A \partial t} \frac{\partial t}{\partial t} + \frac{\partial^2 V(\cdot)}{\partial w_A \partial w_R} \frac{\partial w_R}{\partial t} + \frac{\partial^2 V(\cdot)}{\partial w_A^2} \frac{\partial w_A}{\partial t} = 0 \quad (10)$$



$$\frac{\partial^2 V(\cdot)}{\partial w_R \partial t} \frac{\partial t}{\partial t} + \frac{\partial^2 V(\cdot)}{\partial w_R \partial w_A} \frac{\partial w_A}{\partial t} + \frac{\partial^2 V(\cdot)}{\partial w_R^2} \frac{\partial w_R}{\partial t} = 0 \quad (11)$$

Solving for  $\{\frac{\partial w_A}{\partial t}, \frac{\partial w_R}{\partial t}\}$ , and considering that  $\frac{\partial^2 V(\cdot)}{\partial w_A \partial w_R} = \frac{\partial^2 V(\cdot)}{\partial w_R \partial w_A}$  gives:

$$\frac{\partial w_A}{\partial t} = \frac{\frac{\partial^2 V(\cdot)}{\partial w_A \partial w_R} \frac{\partial V(\cdot)}{\partial w_R \partial t} - \frac{\partial^2 V(\cdot)}{\partial w_R^2} \frac{\partial V(\cdot)}{\partial w_A \partial t}}{\frac{\partial^2 V(\cdot)}{\partial w_A^2} \frac{\partial^2 V(\cdot)}{\partial w_R^2} - (\frac{\partial^2 V(\cdot)}{\partial w_A \partial w_R})^2} \quad (12)$$

$$\frac{\partial w_R}{\partial t} = \frac{\frac{\partial^2 V(\cdot)}{\partial w_A \partial w_R} \frac{\partial V(\cdot)}{\partial w_A \partial t} - \frac{\partial^2 V(\cdot)}{\partial w_A^2} \frac{\partial V(\cdot)}{\partial w_R \partial t}}{\frac{\partial^2 V(\cdot)}{\partial w_A^2} \frac{\partial^2 V(\cdot)}{\partial w_R^2} - (\frac{\partial^2 V(\cdot)}{\partial w_A \partial w_R})^2}, \quad (13)$$

where the denominator is the determinant of the Hessian, and is therefore positive. Define it as  $soc := \frac{\partial^2 V(\cdot)}{\partial w_A^2} \frac{\partial^2 V(\cdot)}{\partial w_R^2} - (\frac{\partial^2 V(\cdot)}{\partial w_A \partial w_R})^2 > 0$ .

Moreover, from equation (8) compute the following derivatives:

$$\frac{\partial^2 V(\cdot)}{\partial w_i \partial t} = -U''(\cdot)t(W - (w_A + w_R)) + U'(\cdot) \quad (14)$$

$$\frac{\partial^2 V(\cdot)}{\partial w_A \partial w_R} = -U''(\cdot)t^2 + \frac{\partial^2 C}{\partial w_A \partial w_R} \quad (15)$$

$$\frac{\partial^2 V(\cdot)}{\partial w_i^2} = -U''(\cdot)t^2 - \frac{\partial^2 C}{\partial w_i^2} \quad (16)$$

Substituting (14), (15), (16), and  $soc$  in equations (12) and (13) gives:

$$\frac{\partial w_A}{\partial t} = \frac{(\frac{\partial^2 C}{\partial w_R^2} - \frac{\partial^2 C}{\partial w_A \partial w_R})(U'(\cdot) + U''(\cdot)t(w_A + w_R - W))}{soc} \quad (17)$$

$$\frac{\partial w_R}{\partial t} = \frac{(\frac{\partial^2 C}{\partial w_A^2} - \frac{\partial^2 C}{\partial w_A \partial w_R})(U'(\cdot) + U''(\cdot)t(w_A + w_R - W))}{soc}, \quad (18)$$

where  $U'(\cdot) + U''(\cdot)t(w_A + w_R - W) > 0$  since  $w_A + w_R - W < 0$ , and standard conditions on the utility function are satisfied such that  $U'(\cdot) > 0, U''(\cdot) < 0$ . Thereby, the sign of  $\frac{\partial w_R}{\partial t}$  and  $\frac{\partial w_A}{\partial t}$  is determined as in equation (9).

## B Appendix: Descriptive statistics

This appendix presents descriptive statistics and denominations of the  $X$  variables included in the estimation of propensity scores and treatment effects. In particular, economic theory and knowledge of the policy setting leads to the

inclusion of 90 control variables. Differently from past literature, this analysis does not ex-ante select among possibly relevant variables. The collected data is richer than that available in previous studies, and the employed random forest approach works well with large and high-dimensional data.

Data errors due to, e.g., typing and misreporting, are analyzed by looking at anomalies in MSW generation over time, specifically, by comparing deviations to the municipal-specific median with the sample upper bound (SUB) defined by  $Q_3 + (Q_3 - Q_1)$  where  $Q$  is the quartile value,  $(Q_3 - Q_1)$  the interquartile range. Municipalities with deviations to the median beyond the SUB are eliminated from the data. Looking at possible outliers, the sample includes all extreme outcome values but excludes units with extremely low values in recycling shares and per capita recycling to increase the comparability of municipalities' waste management practices. Extreme values are eliminated if they are below the sample lower bound (SLB) defined by  $Q_3 - 0.5(Q_3 - Q_1)$  where 0.5 is chosen to avoid that  $Q_3$  takes negative values.

Table 8: Descriptives for treated (T) and never-treated (NT) municipalities (2010-2015)

|            | mean(T) | min(T) | max(T)  | Sd(T)  | (NT)   |        |         |        |
|------------|---------|--------|---------|--------|--------|--------|---------|--------|
| Price      | 0.08    | 0.01   | 0.18    | 0.05   | 0      | 0      | 0       | 0      |
| RW         | 315.41  | 84.79  | 721.51  | 93.14  | 353.94 | 13.13  | 828.52  | 82.84  |
| TW         | 486.00  | 350.62 | 1594.06 | 139.08 | 457.40 | 107.63 | 1193.25 | 127.03 |
| UW         | 170.59  | 35.83  | 1016.42 | 112.62 | 235.47 | 21.62  | 867.26  | 120.53 |
| RWrate     | 0.66    | 0.17   | 0.90    | 0.15   | 0.53   | 0.03   | 0.91    | 0.17   |
| costRW     | 51.30   | 0.11   | 294.66  | 22.39  | 40.34  | 0.11   | 327.44  | 22.48  |
| costUW     | 44.73   | 6.65   | 134.79  | 21.47  | 53.83  | 0.11   | 682.79  | 34.26  |
| costTW     | 96.03   | 21.68  | 301.31  | 26.75  | 94.17  | 2.30   | 805.79  | 38.21  |
| avgCostRW  | 0.17    | 0.00   | 0.60    | 0.07   | 0.18   | 0.001  | 2.24    | 0.12   |
| avgCostUW  | 0.37    | 0.02   | 1.78    | 0.23   | 0.28   | 0.00   | 4.40    | 0.19   |
| avgCostTW  | 0.21    | 0.03   | 0.67    | 0.07   | 0.22   | 0.005  | 1.93    | 0.09   |
| distPayt   | 22.55   | 0.00   | 339.55  | 35.02  | 50.00  | 1.00   | 371.26  | 63.35  |
| distHaz    | 10.71   | 0.00   | 48.13   | 9.92   | 13.80  | 0.00   | 115.40  | 15.40  |
| distInc    | 26.29   | 0.00   | 72.49   | 21.30  | 29.00  | 0.00   | 178.01  | 28.39  |
| distLandf  | 7.70    | 0.00   | 24.51   | 6.49   | 11.21  | 0.00   | 89.93   | 10.95  |
| popDens    | 263.87  | 3.80   | 3581.35 | 396.69 | 337.29 | 1.33   | 7765.52 | 579.97 |
| hhSize     | 2.49    | 1.12   | 6.95    | 0.83   | 2.31   | 1.00   | 7.09    | 0.34   |
| income     | 14.44   | 6.97   | 22.98   | 1.95   | 13.86  | 4.66   | 45.62   | 2.29   |
| migrNet    | 0.00    | -8.80  | 14.68   | 0.55   | 0.00   | -0.14  | 0.18    | 0.01   |
| pop        | 8.52    | 0.12   | 192.84  | 19.27  | 7.35   | 0.03   | 1345.85 | 34.72  |
| foreignPop | 0.10    | 0.01   | 0.19    | 0.04   | 0.08   | 0.00   | 0.41    | 0.04   |
| males      | 0.49    | 0.41   | 0.54    | 0.01   | 0.49   | 0.39   | 0.69    | 0.02   |
| popGrowth  | 0.01    | -0.72  | 0.56    | 0.12   | 0.00   | -1.28  | 1.49    | 0.15   |
| tourism    | 0.36    | 0.00   | 9.34    | 1.01   | 0.33   | 0.00   | 11.64   | 0.85   |
| age0       | 0.05    | 0.03   | 0.06    | 0.01   | 0.05   | 0.03   | 0.06    | 0.01   |

Continued on next page

|              |       |       |       |       |       |       |       |       |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| age14        | 0.14  | 0.07  | 0.21  | 0.02  | 0.13  | 0.00  | 0.35  | 0.03  |
| age65        | 0.22  | 0.11  | 0.34  | 0.04  | 0.35  | 0.05  | 0.51  | 0.05  |
| elemDeg      | 0.31  | 0.19  | 0.42  | 0.04  | 0.31  | 0.14  | 0.57  | 0.05  |
| collegeDeg   | 0.09  | 0.03  | 0.17  | 0.03  | 0.09  | 0.03  | 0.18  | 0.03  |
| rentedHouses | 0.09  | 0.04  | 0.17  | 0.02  | 0.09  | 0.00  | 0.44  | 0.03  |
| hhPerSqMeter | 2.24  | 1.68  | 3.08  | 0.26  | 2.30  | 1.28  | 3.27  | 0.27  |
| oneParentFam | 0.10  | 0.06  | 0.16  | 0.01  | 0.10  | 0.00  | 0.19  | 0.02  |
| students     | 0.06  | 0.04  | 0.08  | 0.01  | 0.06  | 0.05  | 0.09  | 0.01  |
| commuters    | 0.26  | 0.09  | 0.36  | 0.05  | 0.26  | 0.06  | 0.37  | 0.06  |
| deprIndex    | -1.87 | -6.42 | 1.70  | 1.32  | -1.50 | -7.62 | 6.14  | 1.58  |
| outLabRate   | 0.62  | 0.51  | 0.78  | 0.05  | 0.63  | 0.46  | 0.90  | 0.06  |
| unempOutLab  | 0.06  | 0.01  | 0.13  | 0.02  | 0.07  | 0.00  | 0.35  | 0.02  |
| labMarket    | 6.27  | 0.11  | 57.94 | 15.03 | 0.44  | 0.01  | 0.66  | 0.13  |
| polPart      | 0.61  | 0.01  | 0.82  | 0.24  | 0.69  | 0.14  | 0.91  | 0.09  |
| votesBigTent | 0.35  | 0.09  | 0.37  | 0.05  | 0.35  | 0.00  | 0.58  | 0.07  |
| votesLeft    | 0.06  | 0.02  | 0.24  | 0.03  | 0.07  | 0.00  | 0.84  | 0.10  |
| votesRight   | 0.12  | 0.03  | 0.30  | 0.06  | 0.13  | 0.00  | 0.57  | 0.08  |
| localMayor   | 0.85  | 0.00  | 1.00  | 0.36  | 0.76  | 0.00  | 1.00  | 0.42  |
| mayorCenter  | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 0.00  | 1.00  | 0.05  |
| mayorGreen   | 0.03  | 0.00  | 1.00  | 0.16  | 0.01  | 0.00  | 1.00  | 0.11  |
| mayorLeft    | 0.07  | 0.00  | 1.00  | 0.26  | 0.07  | 0.00  | 1.00  | 0.25  |
| mayorOther   | 0.14  | 0.00  | 1.00  | 0.35  | 0.11  | 0.00  | 1.00  | 0.32  |
| mayorReg     | 0.66  | 0.00  | 1.00  | 0.47  | 0.71  | 0.00  | 1.00  | 0.46  |
| mayorRight   | 0.10  | 0.00  | 1.00  | 0.30  | 0.10  | 0.00  | 1.00  | 0.30  |
| mayorAge     | 51.14 | 21.00 | 78.00 | 10.47 | 52.04 | 22.00 | 87.00 | 10.47 |
| yearsOffice  | 1.98  | 0.00  | 5.00  | 1.46  | 1.84  | 0.00  | 8.00  | 1.38  |
| noSeat       | 0.97  | 0.00  | 1.00  | 0.17  | 0.97  | 0.00  | 1.00  | 0.17  |
| provSeat     | 0.03  | 0.00  | 1.00  | 0.16  | 0.01  | 0.00  | 1.00  | 0.10  |
| regionSeat   | 0.01  | 0.00  | 1.00  | 0.07  | 0.00  | 0.00  | 1.00  | 0.06  |
| urban        | 0.48  | 0.00  | 1.00  | 0.50  | 0.38  | 0.00  | 1.00  | 0.49  |
| urbanHigh    | 0.14  | 0.00  | 1.00  | 0.35  | 0.20  | 0.00  | 1.00  | 0.40  |
| urbanLow     | 0.38  | 0.00  | 1.00  | 0.49  | 0.42  | 0.00  | 1.00  | 0.49  |

Table 9: Variables' description. Census indicates 2011 values (ISTAT, 2011).

| Variables' description |   |
|------------------------|---|
| PAYT price             | treatment: unit price on unsorted waste in € per liter                |
| RW                     | recycling waste (RW) per capita (kg)                                  |
| TW                     | total waste (TW) per capita (kg)                                      |
| UW                     | unsorted waste (UW) per capita (kg)                                   |
| RWrate                 | recycling rate (% of total waste)                                     |
| costUW                 | per capita costs of UW management (euros)                             |
| costRW                 | per capita costs of RW management less recycling revenues (euros)     |
| costTW                 | per capita costs of TW management (euros)                             |
| avgCostUW              | average costs of UW management (euros per kg)                         |
| avgCostRW              | average costs of RW management less recycling revenues (euros per kg) |
| avgCostTW              | average costs of TW management (euros per kg)                         |
| distPayt               | distance to closest municipality with PAYT in t-1 (km)                |
| distHaz                | distance to closest hazardous waste treatment facility (km)           |

Continued on next page

|              |  |
|--------------|--|
| distInc      | distance to closest waste incinerator (km)                             |
| distLandf    | distance to closest waste landfill (km)                                |
| popDens      | population density (inhabitants per km <sup>2</sup> )                  |
| hhSize       | average household size (n. household members)                          |
| income       | income per capita (x 1,000 euros)                                      |
| migrNet      | net migrant flow per capita  |
| pop          | population (x 1,000 inhabitants)                                       |
| foreignPop   | share of foreign population  |
| males        | share of male population   |
| popGrowth    | population growth  |
| tourism      | capacity of tourist accommodation per capita (x 1,000)                 |
| age0         | share of population aged less than 5 (census)                          |
| age14        | share of population aged less than 14                                  |
| age65        | share of population aged more than 65                                  |
| elemDeg      | share of population with elementary degree or lower (census)           |
| collegeDeg   | share of population with college degree (census)                       |
| rentedHouses | share of rented houses (census)  |
| hhPerSqMeter | housing density (inhabitants per 100m <sup>2</sup> , census)           |
| oneParentFam | share of single-parent families (census)                               |
| students     | share of population older than 15 and students (census)                |
| commuters    | share of commuters (census)  |
| deprIndex    | social deprivation index (-/+ less/more deprived, census)              |
| outLabRate   | out-of-the-labor force rate (census)                                   |
| unempOutLab  | unemployed and out-of-the-labor force rate (census)                    |
| labMarket    | commuting intensity between municipalities (IIRFL index 0-100, census) |
| polPart      | voter turnout in the 2013 Italian general election (IGR13)             |
| votesBigTent | vote shares for big-tent parties in the IGR13                          |
| votesLeft    | vote shares for extreme left-wing parties in the IGR13                 |
| votesRight   | vote shares for extreme right-wing parties in the IGR13                |
| localMayor   | mayor born in the municipal province (dummy)                           |
| mayorCenter  | centre-party mayor (dummy)   |
| mayorGreen   | green-party mayor (dummy)  |
| mayorLeft    | left-wing mayor (dummy)  |
| mayorOther   | mayor of other party (dummy)   |
| mayorReg     | mayor of local party (dummy)   |
| mayorRight   | right-wing mayor (dummy)   |
| mayorAge     | mayor's age  |
| yearsOffice  | mayor's term of office (years)   |
| noSeat       | no capital (dummy)   |
| provSeat     | provincial capital (dummy)   |
| regionSeat   | regional capital (dummy)   |
| urban        | mediumly urbanized municipality (dummy)                                |
| urbanHigh    | highly urbanized municipality (dummy)                                  |
| urbanLow     | lowly urbanized municipality (dummy)                                   |

---

## C Appendix: The overlap assumption

Overlap analysis indicates that differences in treatment status can be explained to some extent by the included covariates. If  $X$  would fully (not) explain treat-

ment assignment I would observe terminal nodes including only either treated or untreated units and, thus, Generalized Propensity Scores (GPS) equal to one and zero. This would indicate that treatment assignment is deterministic, namely there is no randomness that allows municipalities with identical characteristics to be observed in both states (Heckman *et al.*, 1997). Hence, overlap guarantees that a comparable unit can be found for each municipality. Figure 10 shows that PAYT units are fully in the support of non-PAYT units.

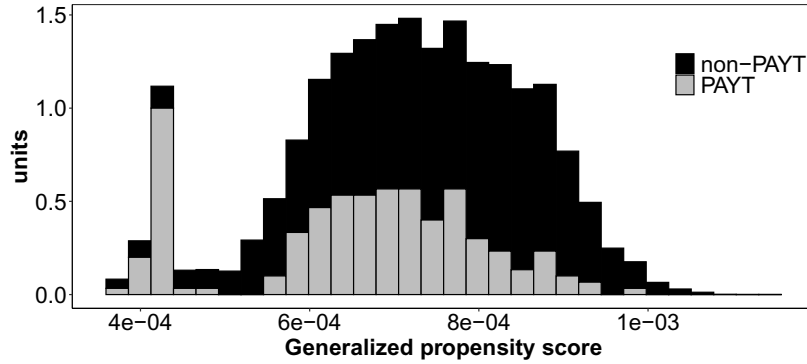


Figure 10: Common support condition for treated (PAYT) and never-treated (non-PAYT).

In particular, treated and control groups are similar in their GPS means, first and third quartiles. Statistics computed following Imbens & Rubin (2015) and Kluve *et al.* (2012) show that the two groups are not apart, namely, the (normalized) difference in estimated GPS is less than one standard deviation (0.62). Also, there is good coverage frequency for both treated and control group meaning that, specifically, 100% (96%) of the treated (control) units have GPS values inside the .025 and .975 quantiles of the GPS distribution of the control (treated) units. Further, all units have close comparisons in the opposite treatment group. In particular, for all treated units and for 96% of the control units there are units with the other treatment status that have differences in GPS less than 10%, a threshold that guarantees unbiased estimates of the causal effects without extrapolation (Imbens & Rubin, 2015). Therefore, causal effects for the control group, and not only for the subpopulation of treated units, can be credibly estimated under unconfoundedness.

## D Appendix

### D.1 Average point-price elasticities

Table 10 shows a comparison with the literature values of average point-price elasticities of waste demands. Typically, the literature relies on the assumption that the mean price and waste quantity are on one point along the linear demand curve (see e.g. Fullerton & Kinnaman, 2000). I derive the slope of the demand curve by estimating the APE in kg, and multiply this value by mean price and waste quantity. Table 10 reports estimates for all waste types and policy years.

Table 10: Average point-price elasticity estimates calculated at mean values with 95% confidence intervals [ci.low; ci.up].

| Year | UW    | ci.low | ci.up | RW   | ci.low | ci.up | TW    | ci.low | ci.up |
|------|-------|--------|-------|------|--------|-------|-------|--------|-------|
| +1   | -0.38 | -0.25  | -0.51 | 0.18 | 0.12   | 0.24  | -0.00 | -0.06  | 0.05  |
| +2   | -0.56 | -0.43  | -0.68 | 0.16 | 0.10   | 0.35  | -0.03 | -0.02  | 0.01  |
| +3   | -0.90 | -0.66  | -1.14 | 0.17 | 0.08   | 0.25  | -0.05 | -0.09  | -0.02 |

The estimate of -0.38 is in line with the literature’s value of -0.344, the average elasticity estimate of UW demands one year after policy computed across 72 studies (Bel & Gradus, 2016). Cross-price elasticity estimates for RW are also in the range of previous estimates: I find higher values than Fullerton & Kinnaman (1996) (0.073), and lower values than Fullerton & Kinnaman (2000) (0.22) and Callan & Thomas (2006) (0.387).

Similarly to the literature, this study consistently estimates price inelastic waste demands. However, comparing elasticities across studies is not straightforward because mean price and waste quantity vary across data sets. The mean price in this study is generally higher, which can explain point-elasticity estimates above the literature’s average.<sup>46</sup> In the first two policy years, my results are in line with e.g. Hong & Adams (1999) who finds that higher prices do not statistically influence total waste on average. However, after three policy years, higher prices lead to significant total waste reductions on average, suggesting that waste avoidance is rather a long-term effect.

<sup>46</sup>For instance, Callan & Thomas (2006) use a mean price of \$0.012 per gallon which translates into less than one cent per liter (vs. mean price of about 8 cents per liter in this study).

## D.2 Testing for effect heterogeneity

Levene’s tests (1960) reject the null hypothesis of no CAPE heterogeneity for all outcomes. Moreover, as suggested in Athey *et al.* (2017), I perform the following heuristic to test for effect heterogeneity. First, I group observations into a high and low APE group using the median CAPE as a threshold. Next, I derive an estimate of the APE for each subgroup by residual-on-residual regression. Results are provided in the Table 11, and show statistically significant differences in APE between subgroups for all outcomes and policy years.

Table 11: Difference in APE (dUW, dRW, dTW in kg) between high and low APE group with 95% confidence intervals [ci.low; ci.up].

| Year | dUW   | ci.low | ci.up | dRW   | ci.low | ci.up | dTW   | ci.low | ci.up |
|------|-------|--------|-------|-------|--------|-------|-------|--------|-------|
| 1    | -2.72 | -2.98  | -2.46 | -5.05 | -5.29  | -4.82 | -5.77 | -6.04  | -5.51 |
| 2    | -3.69 | -3.95  | -3.43 | -5.15 | -5.38  | -4.91 | 12.01 | 11.62  | 12.41 |
| 3    | -4.49 | -4.75  | -4.24 | -6.13 | -6.35  | -5.91 | -3.32 | -3.57  | -3.07 |

## D.3 Price causal effects on waste demands

Figure 11 reports effect heterogeneity across price levels. Price effects on the y-axis (CAPE Kg) are estimated as quantity changes for a one cent price increase. CAPE heterogeneity varies at low versus high prices, *ceteris paribus*: a one cent price increase reduces UW by 10.8 kg at low prices vs. 14.2 kg at high prices, *ceteris paribus*. Higher prices lead to increased recycling by 6.8 kg at low prices vs. 11.6 kg at high prices. The remaining UW amounts are due to significant TW reductions.

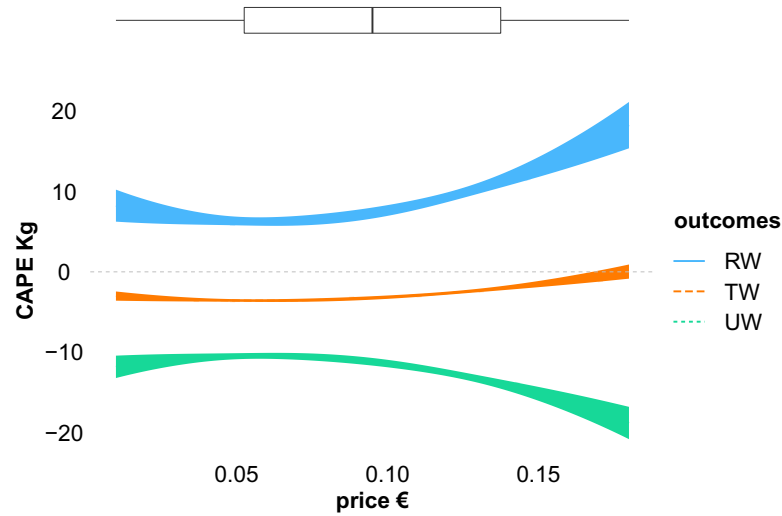


Figure 11: Fitted price effects on waste demands (CAPE Kg) as quantity changes for a one euro cent price increase.

Figure 12 reports effect heterogeneity across pre-policy RW levels. Effect heterogeneity is statistically significant only at high prices: municipalities recycling little before policy have generally higher elasticities, *ceteris paribus*, by on average 1.4 percent points.

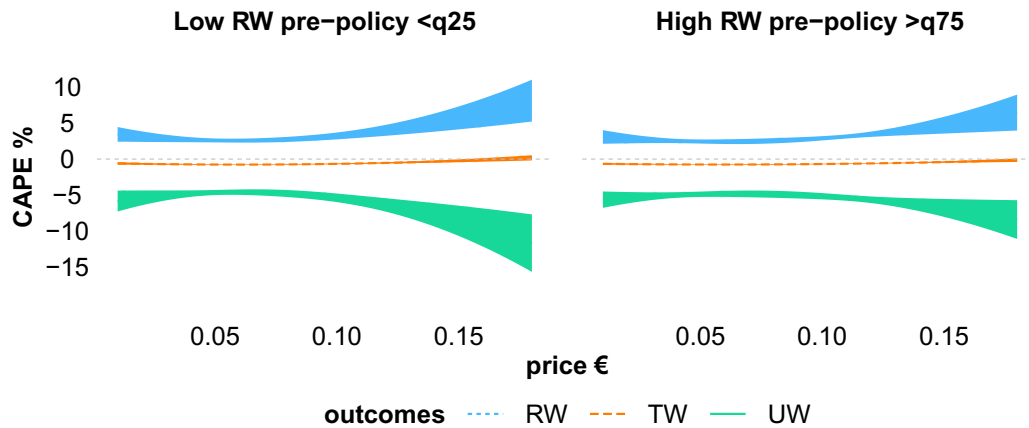


Figure 12: Fitted price semi-elasticities (CAPE) of waste demands by pre-policy recycling levels. Thresholds q25 and q75 indicate first (248 kg) and third (344 kg) quartiles of annual recycling per capita.



## D.4 Heterogeneity in PAYT systems

### D.4.1 Determinants of system choice

Evidence provided by manufacturers of PAYT technology (e.g., magnetic cards, tags and RFID readers) suggests that system choices are driven by three main factors. First, proximity: neighboring municipalities may either share the same waste hauler or influence each other through information dissemination. Second, population density: sparsely populated municipalities are more likely to adopt weight systems (PartItalia, 2020).<sup>47</sup> Third, geography: physical characteristics of the territory such as uphill and one-way roads impact transportation costs and collection modes.

I run a multinomial logit regression in order to assess whether observables can partly explain the choice of the adopted PAYT system. These are: population size (pop), distance to neighboring PAYT municipalities (distPayt), population density (popDens) as described in Table 9 in the Appendix B. Additionally, I distinguished communities by their metro status: urban, peri-urban, semi-peripheral, and very peripheral (the reference category). Categories are defined based on travel times (t) from each municipality to the closest urban center. Table 12 shows the results including all treated observations in the first policy year. The dummy for weight systems is the reference choice.

---

<sup>47</sup>Implementing curbside collection is especially difficult in low-density areas. Also, transportation costs would be high. Thus, communities generally prefer to organize waste collection in centralized disposal areas rather than door-to-door. As volume systems would imply carrying heavy bags and bins, weight systems facilitate households allowing them to carry smaller quantities of waste.

Table 12: Average partial effect estimates of multinomial logit regression. Weight systems are the reference choice. Very peripheral areas ( $t > 75'$ ) are the reference category for geographic dummies.

|   | <i>Dependent variable:</i> |                      |                       |
|---|----------------------------|----------------------|-----------------------|
|   | emptying<br>(1)            | bag<br>(2)           | mixed<br>(3)          |
| pop   | −0.085***<br>(0.032)       | −0.076**<br>(0.036)  | 0.032<br>(0.022)      |
| popDens   | 0.009*<br>(0.005)          | 0.011**<br>(0.005)   | 0.011**<br>(0.005)    |
| distPayt  | −0.006<br>(0.008)          | −0.017*<br>(0.010)   | −0.073<br>(0.046)     |
| peripheral( $40 < t < 75'$ )                              | 1.189*<br>(0.663)          | 1.655***<br>(0.000)  | 28.377***<br>(0.663)  |
| semi-peripheral( $20 < t < 40'$ )                         | −21.713***<br>(0.924)      | 5.599***<br>(0.952)  | 4.797***<br>(0.962)   |
| semi-urban( $t < 20'$ )                                   | −24.078***<br>(0.699)      | 2.523***<br>(0.718)  | 2.132***<br>(0.781)   |
| peri-urban  | −24.819***<br>(1.618)      | 1.111**<br>(0.525)   | −36.060***<br>(0.000) |
| urban   | −20.275***<br>(1.097)      | 4.625***<br>(0.912)  | −5.571***<br>(0.223)  |
| Constant  | 36.339***<br>(1.145)       | 13.313***<br>(0.980) | 0.799<br>(0.706)      |
| Observations  | 194                        | 194                  | 194                   |
| Year and area dummies                                     | Yes                        | Yes                  | Yes                   |
| Akaike Inf. Crit.   | 284.558                    | 284.558              | 284.558               |
| <i>Note:</i> * $p < 0.1$ ; ** $p < 0.05$ ; *** $p < 0.01$ |                            |                      |                       |

Lower population density (popDens) significantly increases the probability to adopt weight programs over any other, as expected. Higher populated municipalities (pop) are more likely to implement weight vs. bag/emptying programs, perhaps to exploit returns to scale. Information dissemination effects (distPayt) positively increase the choice of bag programs. Finally, geographic characteristics seem to have similar effects for bag and mixed program choices with less peripheral municipalities being more likely to adopt these systems (with the exception of “peri-urban” for mixed systems). Most of emptying programs seem to be adopted in peripheral municipalities (rather than very peripheral or urbanized ones).

#### D.4.2 CAPE heterogeneity by system

Table 13: Pairwise comparison of price elasticities (CAPE) on waste amounts by PAYT system: Mean differences (dUW, dRW, dTW in kg per euro cent), standard errors and p-values estimated by Tukey-Kramer’s method (1956).

| PAYT system       | dUW  | s.e. | p-value | dRW   | s.e. | p-value | dTW  | s.e. | p-value |
|-------------------|------|------|---------|-------|------|---------|------|------|---------|
| emptying - bag    | 0.12 | 0.27 | 0.96    | 0.47  | 0.24 | 0.18    | 0.55 | 0.23 | 0.06    |
| mixed - bag       | 0.19 | 0.52 | 0.98    | 0.56  | 0.46 | 0.59    | 0.62 | 0.44 | 0.47    |
| weight - bag      | 0.76 | 0.52 | 0.44    | 0.16  | 0.46 | 0.98    | 1.06 | 0.44 | 0.07    |
| mixed - emptying  | 0.07 | 0.48 | 1.00    | 0.09  | 0.42 | 1.00    | 0.07 | 0.40 | 1.00    |
| weight - emptying | 0.64 | 0.48 | 0.52    | -0.31 | 0.42 | 0.87    | 0.50 | 0.40 | 0.57    |
| weight - mixed    | 0.57 | 0.66 | 0.81    | -0.40 | 0.57 | 0.89    | 0.44 | 0.55 | 0.85    |

Differences are overall statistically insignificant at conventional confidence levels. Estimates only suggest that pay-per-bag systems are (weakly) associated to more waste avoidance (p-values < 0.1): Price semi-elasticities on avoidance are 0.55 and 1.06 kg higher than with pay-per-emptying and weight-based systems, respectively. Overall, my study does not point towards significant differences in price elasticities between PAYT systems.

#### D.4.3 Supporting analyses

*Binary treatment case.*—The binary case assumes that the policy has homogeneous effects across prices, and that policy adoption is homogeneous in prices. Table 14 reports average price effects (APE) estimated as average effects of a binary treatment at the mean price level ( $ATE/\bar{P}$ ). The binary case overestimates APE on UW by 28% on average.

Table 14: Continuous APE vs. Binary APE estimates (kg) at the mean price ( $ATE/\bar{P}$ ) from (level-level) residual-on-residual regression with year clustered standard errors.

| Year | <i>Continuous</i> |      | <i>vs.</i> | <i>Binary</i> |      |
|------|-------------------|------|------------|---------------|------|
|      | UW kg             | s.e. |            | UW kg         | s.e. |
| 1    | -7.38             | 1.30 |            | -7.60         | 2.55 |
| 2    | -8.86             | 1.03 |            | -11.3         | 1.97 |
| 3    | -11.5             | 1.53 |            | -16.5         | 1.08 |

*Difference-in-differences.*—When the timing of treatment varies across units, Difference-in-Differences (DiD) designs are commonly extended to allow for

dynamic average treatment effects by including leads and lags of treatment as regressors (Jacobson *et al.*, 1993). These dummy variables allow policy effects to vary by the number of years relative to separation from policy. Figure 13 presents the results from a dynamic (event-study-like) DiD regression estimated using a standard two-way (unit and time) fixed effects model.<sup>48</sup> The regression includes three leads and lags, and uses the second (unaffected) lead before policy (-2) as a baseline dummy. Since potentially relevant covariates are many and partly collinear, and there is no a priori guidance on which one to exclude, this DiD only controls for time-invariant waste generation determinants captured by municipal fixed effects. Figure 13 plots lead and lagged average price effect (APE) estimates (continuous treatment) with their confidence intervals. Black dots represent statistically significant effects at 5%, and vertical dotted lines indicate the policy adoption year (1).

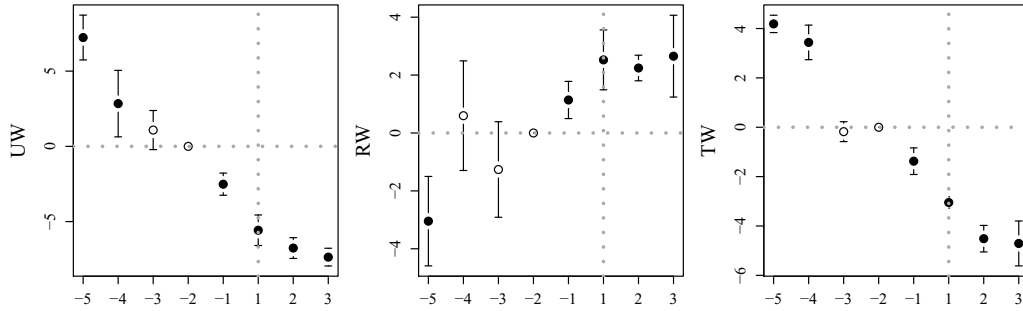


Figure 13: Dynamic DiD estimates (2010-2015,  $n = 19,982$ ) of average price effects (APE kg) on UW, RW, TW with year and unit fixed effects. Clustered standard errors (Driscoll & Kraay, 1994). Statistical significance (5%) indicated by black dots.

Considering the pre-policy period, DiD estimation shows non-zero and statistically significant coefficients on leads. This indicates the presence of pre-trends which invalidate causal effect estimates due to violation of the generalized parallel trends assumption necessary to identify the dynamic DiD estimator (Freyaldenhoven *et al.*, 2019; Abraham & Sun, 2018). The bias possibly derives from having erroneously assumed selection based on time-constant variables. This could be corrected by including time-varying characteristics and allowing variables to have time-varying effects on policy adoption and waste generation. Yet, DiD only allows for time-invariant unobserved effects on the

<sup>48</sup>For the analysis, I use the software R, and, particularly, the package plm (Croissant & Millo, 2008).

outcome, which might not be the case in this setting (Bueno & Valente, 2019; Gobillon & Magnac, 2016). For instance, opportunity costs of waste disposal may have evolving effects over time. Including pre-policy waste outcomes and a high-dimensional set of covariates aims to capture types of heterogeneity that jointly influence the dependent variable and the price. Controlling for this variation is at the heart of this paper’s motivation to employ a forest-based approach.

# The Arab Spring migrant wave and illegal labor on vineyards: Counting the uncountable

with Stefan Seifert\*

*Published in DIW Discussion Papers here.*

**Abstract** About 2,7 million irregular migrants are estimated to work in Europe, of which up to 12% are illegally employed in Italian agriculture. Both farmers and illegal workers have incentives to match on the labor market, and irregular migrant waves may provide additional labor supply shocks. Using the exogenous variation of the 2011 Arab Spring migrant wave on southern Italian coasts and focusing on vineyards, this paper quantifies illegal employment caused by the migration-induced supply shock. Identification relies on abnormal increases in reported labor productivity due to underreported labor hours. Based on farm data at regional level and a dynamic panel model, labor productivity is estimated to abnormally increase by about 11% on average for 2011 and 2012 on vineyards of the landing regions Sicily and Apulia. We show that this corresponds to a total of around 10 million unreported work hours in each year, or 5,500 agricultural work units. We interpret this as an increase in illegal employment with displacement of legal labor due to the migration wave.

**Keywords:** Migration wave, Illegal employment, Labor productivity, Arab Spring

**JEL Codes:** F22, J61, J43

---

\*University of Bonn, Institute for Food and Resource Economics,  
Bonn, Germany. Email: s.seifert@ilr.uni-bonn.de

## Introduction

European industries such as agriculture and construction have long relied on illegal employment and migrant workers to reduce factor costs, and to sustain international competition (Boswell & Straubhaar, 2004). In fact, employing irregular migrants for seasonal, labor-intense jobs is often the only option due to the lack of native labor supply. In this respect, labor supply shocks of irregular migrants may allow an easy match of seasonal labor demand to supply of individuals willing to work illegally.<sup>1</sup> Both irregular migration and illegal work are top priorities in the European Union (EU) due security, financial, and welfare concerns including, e.g., labor market effects on incumbent workers and employers, losses to national revenues, integration of irregular migrants, exploitation of illegal workers, and other violations of labor and human rights (Penal Code, 2009). Policies addressing such concerns need to account for the dimension of the phenomenon, however, the number of irregular migrants and illegal workers is unobserved, and difficult to estimate (Borjas, 2017).

In this paper, we propose a way to identify unreported labor, and we estimate the amount of illegal employment in the grape growing sector caused by a sudden irregular migrant wave. The latter is used as a source of exogenous variation to identify illegal employment under the assumption that an increase in illegal employment leads to an abnormal increase in labor productivity due to underreported work hours. Under plausible assumptions, this paper provides evidence on the displacement of legal by illegal labor in the agricultural sector, which, to our knowledge, is rarely addressed by the literature.<sup>2</sup>

In our empirical analysis, we study labor productivity on vineyards in Italian and French regions, and we analyze the impact of the 2011 Arab Spring migrant wave on illegal employment in the landing regions Sicily and Apulia. The identification of illegal labor uses vineyard labor productivity defined as the ratio of output over labor input. Contrary to its two components, this ratio is rather constant over time<sup>3</sup> and less sensitive to unobserved and time-varying factors such as weather, which allows us to better isolate the effect of the illegal labor

---

<sup>1</sup>We use “irregular” to refer to unauthorized/undocumented persons on the European Union territory, and “illegal” to any worker, regular or not, without legal employment.

<sup>2</sup>This is usually due to the lack of data on illegal labor, and, generally, on the informal sector.

<sup>3</sup>See, e.g., discussions in Jorgenson & Griliches, 1967; Lamouria *et al.*, 1963.

supply shock on vineyards. Thereby, if illegal employment increases after the supply shock, vineyard labor productivity is expected to abnormally increase due to underreported work hours because the output is not impacted by the sudden availability of illegal labor. This is due to the fact that grape growers generally harvest all the output. As a result, an abnormal increase in labor productivity after the shock indicates displacement of legal with illegal employment. This, however, does not necessarily indicate that incoming irregular migrants displace incumbent workers. Since identification only requires that illegal labor can substitute unskilled labor flexibly without sacrificing learning effects, former legally employed workers might become illegally employed due to increased competition.

We use the exogenous variation from the 2011 Arab Spring migrant wave that hit the southern Italian regions of Apulia, Calabria and, primarily, Sicily. This was the largest wave of the last decades crossing the central Mediterranean with about 64,000 detected and further undetected landings (FRONTEX, 2016; INEA, 2014). In particular, in the first half of 2011 about 50,000 migrants landed illegally, compared to an average of about 4,000 persons in the same period in the two prior years. This labor supply shock can be considered as exogenous because neither Arab Spring migrants nor southern Italian coasts self-selected into treatment, and labor markets of origin- and host-countries are not interdependent. Indeed, the migrant wave was so unexpected that Italy declared the state of emergency already in February 2011. Specifically, Italy lacked resources for continuous coastal monitoring as well as migrants' first aid, identification, and detention (Gola, 2015).

In our analysis, the “treatment” is the migration-induced supply shock of illegal labor, and the “treated” regions are Sicily and Apulia<sup>4</sup> which are the only regions where large numbers of incoming, undetected migrants may have supplied illegal labor to local farmers – in our case, to grape growers.<sup>5</sup> As a control group we use Italian and French regions where migrant landings, namely, irregular migrant shocks, do not take place. We find it plausible that irregularly landed migrants may travel to control regions, though not in masses.

---

<sup>4</sup>Calabria, a region located between Sicily and Apulia that likely received a large number of migrants as well, is excluded due to missing data, but it also has only a very small grape sector.

<sup>5</sup>Background information on illegal labor channels for undetected migrants in the treated regions is provided in Section 1.1.



In fact, travelling is difficult due to the lack of documents and the risk of immediate expulsion if caught. Therefore, we assume that an irregular migrant shock occurs only in the landing regions, and not in other Italian and French regions.<sup>6</sup>

The illegal labor supply shock is expected to be absorbed by the grape growing sectors of Sicily and Apulia due to the large demand for seasonal, cheap labor as well as sizable informal agricultural labor markets. The demand side is characterized by a large number of on average very small farms (80% with less than 5 hectares) that are, however, worldwide leaders in table grape production, and also produce large quantities of wine, mostly low-quality (INEA, 2012). Since the 2000s, increasing international competition, shrinking demand globally for wine, climatic shocks, and failed agricultural policies leading to deregulation of the EU wine sector (CMO, 2008) required grape growers - and in particular, (small) less favored vineyards - to reduce factor costs (Gaeta & Corsinovi, 2014). One option is to rely on illegal and migrant workforce. In particular, in Sicily (Apulia), there are about 114,000 (110,000) agricultural workers of which about 32,000 (26,000) foreigners (non-EU 15) employed mostly in tree crops (over 30%) with seasonal contracts (98%), fully legal for only 30%, and with 90% of the wages below the minimum admitted by law (INEA, 2012).

Farm migrant labor in Sicily and Apulia is often absorbed through *caporalato*, an increasingly widespread informal system to recruit and exploit underpaid workforce through intermediaries (Flai-Cgil, 2016).<sup>7</sup> There is considerable anecdotal evidence that shows the link between migration waves and illegal farm work: For example, in August 2011, the first self-organized revolt of farm pickers took place, leading to investigations that revealed illegal migrant trafficking between Apulia, Calabria, and Sicily (Spagnolo, 2017). However, there is little empirical evidence on the amount of illegal labor in agriculture and its effects on the agricultural labor markets, especially in the European Union

---

<sup>6</sup>We cannot extend our sample to regions of other European countries because they are either recipients of large migrant waves, or have very different viticultural traditions, grape sectors, climatic conditions, and regulations of agricultural labor markets (Section 2). Sensitivity of our results is tested against different choices of the control group (Section 4).

<sup>7</sup>Over 400,000 irregular workers, most often undocumented migrants, are estimated to be part of such system, for a 1.8 billion euro loss to national revenues each year (Flai-Cgil, 2018).

(EU). This paper contributes to filling the gap.<sup>8</sup>

For estimation, we use farm-level data aggregated at the regional level for Italy and France between 1999 and 2012 supplied by the Farm Accountancy Data Network (EC, 2017). We analyze causal effects using a Difference-in-Differences (DiD) framework in a dynamic linear panel regression model. Thereby, using a lagged dependent variable as an exogenous predictor accounts for the persistence that characterizes vineyard labor productivity while simultaneously controlling for time-varying effects of unobserved heterogeneity. Various tests support this model, and indicate that fixed effect approaches are likely not suited to account for the underlying dynamics in the data. As a robustness check against the residual presence of unobserved fixed (time-invariant) effects in the dynamic model specification, we implement an Anderson-Hsiao type regression (Anderson & Hsiao, 1981). To further validate our results under different choices of the control group, we use  $k$  nearest neighbor matching to identify the best control matches for each treated region, and we re-estimate the DiD model for different values of  $k$ . Finally, we check for model misspecification and allow for more flexible functional forms, and we estimate the causal effect on the selected model using the post-lasso estimator as introduced by Belloni *et al.* (2012, 2013).

Our results point to an increase of illegal employment on vineyards after the 2011 migrant wave. Indeed, we find that this labor supply shock has a statistically significant average causal effect on labor productivity on Sicily's and Apulia's vineyards, increasing by around 11% on average over the post-treatment period (2011 and 2012). Under further assumptions, we show that this effect corresponds to around 10 million unreported hours in the treated regions in each year – or around 5,500 unreported agricultural work units (each defined as working 1800 hours a year; EC, 2017). In other terms, the migration-induced supply shock caused the average farm to illegally employ about one full-time grape picker for its harvest (about 30 days for one grape variety). As mentioned before, these workers can be undocumented migrants from the 2011 migration wave, and/or other workers available to work illegally, including former legally employed workers. Further, due to high labor intensity and low mechanization of the grape growing sector of the treated regions, a substitution

---

<sup>8</sup>More about the existing literature on the identification of illegal labor, and its effect on legal employment is discussed in Section 1.

effect of capital with illegal labor is rather unlikely. Given the absence of any technological shock in grape harvesting, price shock of the harvested grapes, or additional labor market shocks in the analyzed period, our findings are likely due to the displacement of legal with illegal workforce caused by the 2011 inflow of migrants.

It should be noted that our estimates of the treatment effect on reported labor productivity is only a lower bound of the actual treatment effect. The reason for this is that reported labor productivity is already upward biased before the treatment due to the employment of illegal labor. Thus, also estimates of the displacement of legal workforce is only a lower bound. Moreover, the treatment could have also caused displacement of former illegal labor with incoming illegal labor which cannot be captured by our identification strategy.

The remainder of the paper is organized as follows: Section 1 provides further background information about this study and summarizes related literature. Section 2 outlines our empirical strategy. The data are described in Section 3. Section 4 presents the results, and Section 5 concludes.

## **1 Background and Literature**

### **1.1 Farm labor and the 2011 migrant wave**

The Italian Association for Juridical Studies on Immigration (ASGI, 2015) reports numerous violations of the EU Directive on illegal immigration and illegal employment of migrants without the required legal status in the EU (2009/52/EC). In Italy, estimates report over 400,000 illegal farm workers of which around 80% are migrants, often undocumented (Assosomm, 2016; Flai-Cgil, 2014). Indeed, about 30% of total agricultural employment in Italy, and up to 70% in Sicily and Apulia at local level, is illegal (Flai-Cgil, 2012). These estimates are in strong contrast to the only 5,200 agricultural labor contracts found to be illegal by the Labor Inspectorate in 2018 (Flai-Cgil, 2018). In addition, illegal workers in agriculture seem to be increasing over time, with up to 50,000 more from 2016 to 2018, producing an overall value of 4.8 billion euros and causing losses to national revenues for around 1.8 billion euros (Flai-

Cgil, 2018).<sup>9</sup> Often, illegal migrant labor is exploited with wages below the legal minimum thresholds of around 850 euros per month (about 5 euros per hour), averaging 40% lower than domestic wages (ETI, 2015). In fact, studies report wages between 1.60 and 3 euros per hour over a 12 to 16-hour working day, with which workers have to pay intermediaries for transportation, food, and accommodation (Palmisano & Sagnet, 2016).

The 2011 migrant wave might have further contributed to increase illegal employment in agriculture. In the following, we outline (I) why and (II) how Arab Spring migrants avoid or leave the legal asylum framework, and (III) provide evidence that only Sicily and Apulia are the recipients of the illegal labor supply shock.

(I) Landed migrants want to avoid expulsion and, most often, reach other EU destinations. However, high risks of asylum rejection which implies detention and expulsion, or even immediate repatriation are strong incentives to avoid or leave first aid and reception centers in the landing regions. In 2011, the risk of asylum rejection was very high for most migrants because local administrations considered many countries of origin to be safe (CeSPI, 2012).<sup>10</sup> Also, an immediate repatriation agreement was signed with Tunisia in April 2011 (Il Post, 2011).<sup>11</sup> Lastly, migrants' living conditions in overcrowded temporary facilities were usually very bad, and law infringements including violations of human rights were often reported (e.g., in ASGI, 2011). Thus, migrants have strong incentives to find alternative ways to make money and to obtain (forged) documents in the landing regions. One common way is to use the intermediation of gangmasters, called *caporali*, who negotiate with farmers and supply workers (Spagnolo, 2017; Flai-Cgil, 2012).

(II) Once landed, migrants can enter illegal labor channels in several ways. Many migrants land on Italian shores undetected (INEA, 2014). Detected migrants obtain first aid in emergency shelters called hotspots. In 2011, of the approximately 64,000 detected landings on southern Italian shores, only around

---

<sup>9</sup>Compared to the overall value of the informal economy (200 billion euros), the value of unreported labor amounts to 77 billion euros (4.8% of GDP), and comes from about 3 million workers (ISTAT, 2017).

<sup>10</sup>Indeed, 65% of political asylum requests were rejected, though, most verdicts were appealed (MPP, 2012; Sasso & Sironi, 2012).

<sup>11</sup>Of the 19,000 Tunisians held in temporary camps for expulsion, only about 4,000 and 2,400 were actually expelled in 2011 and 2012, respectively (ISMU, 2015). This led to revolts and further fleeings.

55,000 arrivals are recorded at the hotspots (SPRAR, 2011). Further 11,700 migrants seem to have disappeared in the initial months of 2011 (Polchi, 2011). Yet, without documents and money, migrants are most often not able to travel to their destination country immediately. Among the landed migrants, those not applying for asylum are sent to detention and expulsion centers. Instead, asylum seekers are hosted in reception camps called Centers for First Assistance (abbreviated as CDA), and Centers for Assistance of Asylum Seekers (CARA). In these camps, authorities decide whether immigrants should be expelled. Migrants stay in such facilities also for years, and, despite surveillance, many cases of migrant fleeing camps, temporarily or permanently, have been reported (Gatti, 2015).

(III) As a consequence of the 2011 migration wave, the number of migrants hosted in CDA and CARA tripled: From around 12,000 and 11,000 in 2009 and 2010, respectively, to around 32,000 in 2011 (SPRAR, 2012; ASGI, 2012). 95% of migrants in CDA and CARA are hosted in Sicily, Apulia, and Calabria (SPRAR, 2012). In particular, in 2011, Sicily's and Apulia's centers hosted around 10,800 and 13,800 migrants, respectively, each hosting around 10,000 migrants more than in 2010, and with a 15 and 10 percentage point increase of the share of hosted migrants, respectively. On the contrary, centers in other Italian regions, specifically, Lazio, Marche, and Friuli Venezia Giulia on average reduced both the number and the share of hosted migrants from around 2,100 in 2010 to 1,600 in 2011, and from around 18% in 2010 to 5% in 2011. Therefore, we may expect that, despite surveillance, only migrants hosted in CARAs and CDAs of Sicily, Apulia, and Calabria may leave camps in large numbers, and work in the informal sector on a regular basis.

Furthermore, a number of asylum seekers, only after completed identification procedures, are relocated in reception centers across the country within the System for Protection of Asylum Seekers and Refugees (SPRAR). Numbers reveal that Italian regions did not experience abnormal arrivals or departures of asylum seekers in their SPRAR centers over the years – which may have caused shocks to the local formal and/or informal agricultural labor market (see Table 7 in the Appendix). Moreover, with funds available for emergency management, the Italian government provided about 22,000 extra places in hotels and apartments to host asylum seekers, which were distributed in each

region according to its population (DPC, 2017).<sup>12</sup> As a result, each region received a relatively low number of migrants (around 1,000 on average). Regional numbers as of September 2012 as well as the share of the 17,859 people hosted in each region are reported in Table 1.<sup>13</sup>

Table 1: Number of Arab Spring asylum seekers and refugees relocated across Italian regions in apartments and hotels (source: Sasso & Sironi, 2012).

|                   | n.people | share | pop.* |                   | N.people | Share | Pop.* |
|-------------------|----------|-------|-------|-------------------|----------|-------|-------|
| <b>North-East</b> |          |       |       | <b>North-West</b> |          |       |       |
| Friuli V.G.       | 397      | 0.02  | 1.2   | Aosta V.          | 20       | 0.00  | 0.1   |
| Veneto            | 1274     | 0.07  | 4.9   | Piedmont          | 1549     | 0.09  | 4.4   |
| Trentino A.A.     | 172      | 0.01  | 1.1   | Lombardy          | 2548     | 0.14  | 10    |
| Emilia R.         | 1585     | 0.09  | 4.5   | Liguria           | 540      | 0.03  | 1.6   |
| <b>Center</b>     |          |       |       | <b>South</b>      |          |       |       |
| Tuscany           | 1141     | 0.06  | 3.7   | Abruzzo           | 11       | 0.00  | 1.3   |
| Marche            | 462      | 0.03  | 1.5   | Molise            | 116      | 0.01  | 0.3   |
| Lazio             | 1790     | 0.10  | 5.9   | Campania          | 2155     | 0.12  | 5.8   |
| Umbria            | 338      | 0.02  | 0.9   | Basilicata        | 200      | 0.01  | 0.6   |
| <b>Islands</b>    |          |       |       | Apulia            | 1071     | 0.06  | 4.1   |
| Sicily            | 1110     | 0.06  | 5.1   | Calabria          | 956      | 0.05  | 2.0   |
| Sardinia          | 424      | 0.02  | 1.6   | <b>Total</b>      | 17.859   |       | 60.5  |

\* total regional population in million people

As in SPRAR facilities, migrants could not take an unjustified daily leave and were not allowed to work.<sup>14</sup> Further, in addition to board and lodging, they also received pocket money and temporary documents to access health care services (ANCI, 2011). Therefore, it does not seem plausible that relocated migrants illegally worked on vineyards in other parts of Italy. In fact, both asylum seekers and refugees had no incentive to put their status at risk by working illegally. Also, most refugees seem to have left Italy after positive asylum decisions (Labanca, 2016). However, as of November 2012, most of the pending and rejected asylum seekers did not leave migrant facilities waiting for asylum decisions and appeals, respectively (Lambruschi, 2012). Thus, the

<sup>12</sup>I.e., 10,000 migrants for 100,000 inhabitants. Also note that, in 2011, Lombardy's SPRAR hosted less asylum seekers (see Tab. 7), but the highest share of the extra places was assigned to this region.

<sup>13</sup>Note that 22,216 persons were hosted as of November 2011, therefore regional numbers were slightly higher compared to September 2012.

<sup>14</sup>Officially, asylum seekers cannot work for six months from the asylum application, however, practically, work permits are very difficult to obtain also within 12 months from the request (MPP, 2013).

Italian government offered an accelerated procedure to regularize asylum applicants through concession of an humanitarian visa (Interior Ministry, 2012). Thereby, from 2013 onwards, rejected asylum seekers, in particular, may have entered illegal labor channels to avoid expulsion (Giangrande, 2017).<sup>15</sup>

## 1.2 Literature

A vast literature on the impacts of immigration on productivity and labor market outcomes has developed (for overviews see Peri, 2016; Dustmann *et al.*, 2016; Okkerse, 2008). This includes studies on the long-term consequences of immigration as well as on the labor market effects of sudden migration waves. However, the effects on the informal sector have been rarely addressed.

This paper, by proposing a new way to identify unreported labor and quantifying illegal employment causal to an irregular migrant wave, contributes at the intersection of three streams of literature. The first stream aims to identify and quantify irregular migrants already present in the host country, and studies their labor supply (see, e.g., Borjas, 2017; Warren & Passel, 1987; Kelly, 1977). The second evaluates the effects of immigrants working illegally on legal employment in the agricultural sector (see, e.g., Venturini & Villosio, 2008; Vaiou & Hadjimichalis, 1997). Finally, the third analyzes the effects of migrant labor supply shocks on illegal employment and, generally, on the informal labor market using available survey data on informal workers and irregular migrants (see, e.g., Tumen, 2016; Ceritoglu *et al.*, 2015; Del Carpio & Wagner, 2015).

Borjas (2017) represents the latest example within the first stream of literature. In his paper, Borjas extends past methods to identify undocumented migrants at individual level in survey data based on residual calculations and reconstructions of foreign-born persons that cannot be labelled as legal. This allows the analysis of long-term trends of migrant illegal employment, and the comparison with legal migrant as well as native employment trends. Findings show that irregular male migrants participate more in the labor force than natives and legal migrants, a gap found to increase over time. We relate to this literature in scope as we try to “count the uncountable” by estimating the population of illegal workers when an illegal labor supply shock occurs, and by giving an indication of the magnitude of the phenomenon. Also, we contribute

---

<sup>15</sup>For this reason we do not extend our empirical analysis after 2012.

by proposing a new way to estimate the relative size of the illegal labor employed when a source of exogenous variation is available but data on incoming as well as incumbent populations are not.

In the second stream of literature, to the authors' knowledge, Venturini (1999) is the study most closely related to ours. It uses national-level estimates of the number of illegal immigrant workers in Italy to estimate substitution and complement effects with legal workers. Results reveal that the former compete with the latter in the agricultural sector due to the unskilled, homogeneous nature of farm labor, causing legal labor to be displaced. Similarly for Greece, Vaiou & Hadjimichalis (1997) and Lianos *et al.* (1996) find that illegal employment of agricultural labor competes with legal labor, lowers factor costs, and has a dampening effect on wages of legal workers. We contribute to this literature by providing further evidence of the displacement effects of inflows of irregular migrants on legal workers in the agricultural sector for the Arab Spring migrant wave.

Finally, in the third stream of literature, a series of papers analyzes the labor market impacts of the migration waves following the outbreak of the Syrian civil war in 2011. Focusing on Turkey, Ceritoglu *et al.* (2015) estimate negative causal effects on employment of informal native workers using survey data in a DiD framework. Likewise, Del Carpio & Wagner (2015), Tumen (2016), and Balkan & Tumen (2016) conclude that a strong displacement of natives by immigrants occurs especially in the informal sector. This literature concludes that the combination of prevalent informal employment along with a supply shock of undocumented refugees increased the magnitude of the negative effects of the Syrian migrant wave on natives' employment. However, Peri (2016) argues that the estimation of these causal effects could be biased due to potential war spillovers between the neighboring Syria and Turkey not directly related to the migrant influx, preventing potential outcomes caused by forced migration to be disentangled from other labor market adjustments. The question of which type of worker is actually displaced by incoming irregular migrants is very important also in our case, though, it remains open because we lack statistical information to estimate this relation. However, the above conclusions are in line with our findings which suggest that, due to the Arab Spring migrant wave, legal labor was partly displaced by illegal labor on vineyards – a sector strongly



characterized by informality, especially in southern Italy.<sup>16</sup>

## 2 Methodology

### 2.1 Identification and estimation

We aim at identifying the causal effects of the migration wave on illegal employment by analyzing labor productivity under the assumption that employment of illegal labor leads to underreported labor input and, thus, overreported labor productivity. In the agricultural sector, unskilled labor productivity is fairly constant over time once we control for farm inputs, and production environment, e.g., weather conditions (Jorgenson & Griliches, 1967; Lamouria *et al.*, 1963). This persistence of outcomes may be due to time-invariant, unit-specific unobservables (fixed effects) and due to so-called state dependence. State dependence (Heckman, 1981), i.e., a dynamic outcome process in which past outcomes help predicting future outcomes, can be explained by observed past behavior, but also by time-varying effects of unobserved heterogeneity. If ignored, the latter would cause inconsistent model estimates due to Omitted Variable Bias (OVV).

In particular, in our setting, the estimation of labor productivity would suffer from OVB if regional differences in farmers' propensity to hire illegal labor are not accounted for. This varies among regions and over time, and its determinants are, e.g., organized crime intensity, and farmers' attitudes toward risk and law obedience due to different incentive and monitoring systems at regional level. Further, the propensity to hire illegal labor in a region is likely serially correlated because farmers who have hired illegal labor in the past are more likely to do so also in the future.

To account for these effects, we model labor productivity as a dynamic linear process with limited memory under the assumption of sequential exogeneity. The latter means that the inclusion of a Lagged Dependent Variable (LDV)

---

<sup>16</sup>Another paper closely related to ours estimating the short-term impact of the Arab Spring migrant wave is provided by Labanca (2016). Analyzing selected subgroups of (legal) refugees, and based in particular on Tunisians, results indicate that the migration wave induced natives to shift sectors of employment, specifically from mining, wholesale trade, restaurants and hotels to construction. Interestingly, no effects on the legal agricultural sectors are found.

makes the outcome conditionally independent on past values of observed and unobserved variables. In particular, the inclusion of a LDV controls for state dependence as long as it absorbs time-varying effects of unobserved heterogeneity at regional level. On the contrary, a static model with fixed effects is not able to capture state dependence. While combining both LDV and fixed effects is possible, testing for unobserved time-invariant, unit-specific effects allows to identify the most likely source(s) of this persistence (Breusch & Pagan, 1980; Honda, 1985).

The model to identify the average causal effect of the illegal labor supply shock writes:

$$y_{it} = \rho y_{it-1} + \gamma X_{it} + \delta D_{it} + \mu_i + \varepsilon_{it}, \quad (1)$$

where subscripts  $i$  and  $t$  indicate units and time, respectively,  $y_{it}$  is labor productivity,  $y_{it-1}$  is the (observed) LDV capturing (unobserved) time-varying, unit-specific effects on  $y_{it}$  through the parameter  $\rho$ ,<sup>17</sup>  $X_{it}$  are other exogenous regressors,  $\mu_i$  are (unobserved) fixed effects,  $\varepsilon_{it}$  error terms, and  $D_{it}$  is a treatment dummy. Following the potential outcome approach by Rubin (1974), the DiD estimator  $\hat{\delta}$  estimates the Average Treatment effect on the Treated (ATT) as the difference of two differences: the average outcome in the treatment group, i.e., Sicily and Apulia, before and after shock, and the average outcome in the control group before and after shock. In particular, the ATT is the state-dependent causal effect of the migration-induced labor supply shock on labor productivity. This is a conditional DiD estimator (Fitzenberger *et al.*, 2009) where the treatment effect is conditional on the reported labor productivity in the previous period  $t - 1$ , including the amount of misreporting already in  $t - 1$ . Thus, by conditioning on the reported labor productivity in the previous year, we account for the short-run dynamics in labor misreporting.

If outcome persistence is mainly due to state dependence, the model includes only the dynamic component, and  $\mu_i$  drops. Excluding  $\mu_i$ , the dynamic model (1) can be estimated by pooled OLS (DPOLS) under the assumption of exogenous regressors and sequential exogeneity. If the LDV absorbs residuals' autocorrelation, parameter estimates are unbiased and consistent.<sup>18</sup>

---

<sup>17</sup>Any LDV model assumes two (tested) stability conditions on the autoregressive process, i.e., covariance stationarity ( $|\rho| < 1$ ) and weak dependence of the outcome (Hsiao, 2014).

<sup>18</sup>If, instead, autocorrelation would still be present, the model is likely non-stationary even if  $|\rho| < 1$ , causing inconsistent and biased estimates (Keele & Kelly, 2006), and, in particular,

The identification of  $\delta$  in (1) is based on two assumptions: conditional independence (CIA), and no externalities of treatment aka stable unit treatment value assumption (SUTVA).<sup>19</sup>

The CIA requires that, conditional on explanatory variables, the assignment of the treatment is as good as random. Indeed, the 2011 migration wave was an exogenous shock as it was unexpected and abnormally large, only hitting the southern Italian coasts due to their geographic vicinity. In particular, migrants escaping from wars did not select southern Italy as their preferred destination due to its large agricultural informal sector. Because neither migrants nor the treated regions self selected into treatment and because regressors are chosen to be exogenous, we consider this assumption as fulfilled. In addition, political and economic dynamics of the immigrants' countries of origin do not affect southern Italy as these economies are not interconnected and no spillovers occur.

Lastly, SUTVA states that no spillover effects on the control units take place after the shock. Indeed, no treatment effect is expected outside Sicily and Apulia (and excluding Calabria) for the years 2011 and 2012. In particular, migrants either avoiding or leaving first aid and reception camps are irregular and at risk of expulsion; thus, provided that they find informal employment in the landing regions, they have no incentive to leave, travel undocumented across Italy, and run a higher risk of detention and expulsion. In addition, relocated migrants across Italy do not represent a supply shock to other regional informal sectors. The main reason is that asylum seekers and refugees had no incentive to put their status at risk by working illegally (see related discussion in Section 1.1). However, with the end of the state of emergency on January 2013, the Italian government interrupted the extra financing of asylum seeker and refugee support services, which may have caused rejected asylum seekers to stay irregular and to look for an irregular job (Giangrande, 2017). For this reason we do not extend our analysis after 2012.

In any observational study, the choice of the optimal control group is *ex ante* difficult. In our case, a panel instead of a pure time series regression allows to distinguish the treatment from a year effect such as general weather conditions. However, control regions should be untreated and resemble the treated regions

---

upward bias for  $\hat{\rho}$ , and downward bias for the other coefficients (Achen, 2000).

<sup>19</sup>Note that parallel trends assumption in the case of a LDV model is generally not required to hold as there is regression to the mean.

in the relevant observed characteristics. Thus, we use Italian and French regions as the control group, but we cannot extend our sample to regions of other European countries because they are either recipients of large migrant waves, or have very different viticultural traditions, grape sectors, climatic conditions, and regulations of agricultural labor markets. On the contrary, Italian and French regions are comparable in many respects and are therefore suitable as a control group. Specifically, France and Italy share a border and have similar climatic conditions with warm Mediterranean climate in the south and temperate oceanic climate in the north. Further, both countries are EU member states, grape growers are working under mostly identical regulation, and both countries have a long grape growing tradition. Finally, also on French farmlands, migrant seasonal workers are or become often illegal: Most workers are recruited from abroad via foreign recruitment agencies, and are given a temporary work permit for the harvesting season after which some of them are often offered to work illegally (Flai-Cgil, 2016).<sup>20</sup> Also, farmers face limited risks in terms of sanctions due to lack of labor inspectorate effectively monitoring working conditions and paid wages (Meyer & Dumortier, 2014).<sup>21</sup>

Furthermore, since no treatment is expected to occur before the migrant wave, the assumption of joint significance of treatment dummies before treatment, aka in-time placebos, should be tested, and rejected (Autor, 2003).

## 2.2 Robustness checks

As a robustness check against the residual presence of unobserved fixed effects,  $\mu_i$ , we implement an Anderson-Hsiao type regression by first differencing the model and instrumenting the endogenous first-differenced LDV with lagged outcomes (AH, Anderson & Hsiao, 1981). First differencing may, however, induce new problems: First, if the DPOLS errors in (1) are not serially correlated, this induces autocorrelation in the first-differenced errors, weakening the exogeneity of the chosen lagged outcomes as instruments; second, first differencing may induce outcome cross-sectional dependence that was not present

---

<sup>20</sup>In 2002, the increasing magnitude of this phenomenon and the need to document illegal practices led to the creation of the Association for the Defense of Foreign Seasonal Workers in Agriculture aka CODETRAS (see [codetras.org](http://codetras.org)).

<sup>21</sup>To address this problem, the 2014 National Action Plan against Human Trafficking reformed and extended the mandate of labor inspectors (NAPHT, 2014).

in levels.

In addition, we perform a sensitivity analysis on the choice of the control group. Given that some regions are more likely to be similar to the treated regions in terms of pre-treatment characteristics, we use  $k$  nearest neighbor matching to identify for each treated region the  $k$  best control matches based on covariate balance. Successively, we re-estimate the DiD model for different values of  $k$ .

To check for model misspecification and to allow for more flexible functional forms, we use the post-lasso regression method as introduced by Belloni *et al.* (2012, 2013). With this approach, in a first stage, a model is selected using a lasso-type regression applied to a high-dimensional model. In a second stage, the ATT is estimated using the selected model (post-lasso). For this purpose, we first extend our initial model by other farm-related variables and many technical variables from the chosen covariates, such as interaction terms, log specifications, as well as second- and third-order orthogonal polynomials. We obtain a high-dimensional model with many parameters relative to the sample size that can be estimated under approximate sparsity and, as before, conditional sequential exogeneity. Lasso regularizes the regression by the penalized L1-norm to avoid overfitting, thereby selecting the variables with the best explanatory power (for details see, e.g., Chernozhukov *et al.*, 2013; Chernozhukov *et al.*, 2017). In the second stage, the selected model is used to estimate the ATT. This post-lasso regression undoes the regularization bias caused by the parameter shrinkage.<sup>22</sup> By doing so, we retain those variables that can increase prediction accuracy while reducing variance. However, model specifications selected with a lasso approach lack theoretical justification, and the economic interpretation of parameter estimates is often not straightforward. For this reason, we use post-lasso as a complementary approach to check if our ATT estimates are robust to alternative model specifications.

### 3 Data

We use data from the Farm Accountancy Data Network (FADN, EC, 2017) and explanatory variables from the Eurostat labor force survey (LFS, EUROSTAT,

---

<sup>22</sup>Post-lasso is shown to perform at least as well as lasso in terms of the rate of convergence, and has the advantage of a smaller regularization bias (Belloni & Chernozhukov, 2013).

2017). The dataset is a balanced panel over the period 1999-2012. Each cross-section is the average farm of each region on the *NUTS 2* level. The sample is representative due to stratified sampling and weighting. The farms in our sample are vineyards, i.e., farms specialized in grape and wine production.<sup>23</sup> The sample consists of 25 regions of which 14 are located in Italy and 11 in France, adding up to in total 350 observations.<sup>24</sup>

We consider the two southern Italian regions Sicily and Apulia as the treated units, for which the treatment, i.e., the illegal labor supply shock, takes place in 2011, and may have effects in 2012. As Figure 1 highlights, a sudden increase in landings of migrants on the southern Italian coast took place in the spring preceding the grape harvest of that year. Further, as previously discussed, the number of landings is likely underestimated (INEA, 2014), as well as a large number of irregular migrants seem to have avoided or fled migrant shelters, and might have been available as labor force from 2011 onwards.

However, despite data availability for later years, we restrict the analysis to 2011 and 2012. The reasons are twofold. First, the effect of the treatment could be confounded by the additional landings registered in 2013 (see Figure 1 and FRONTEX, 2016). Second, potential spillover effects from 2013 onwards may violate the assumption on the untreated status of the control regions (for a discussion, see Section 2.1).

To analyze labor productivity, we define our logged dependent variable *LabProd* as the total output from crops in euros divided by total hours worked, which is the sum of all paid and unpaid hours worked and includes family labor. While this measure controls effectively for seasonal/unseasonal jobs and working regimes, it should be noted that it may vary due to quantity and price variations, with the latter being the main source of concern. However, we are confident in ruling out such effects for several reasons. First, price effects could be sizable only for high-quality wines, i.e., those with Protected Designation of Origin (PDO). In 2011, in Sicily (Apulia) PDOs account for only 4% (15%) of the regional wine production, and for only 1% (6%) of total Italian PDO

---

<sup>23</sup>Additional farm income from other agricultural activities plays a minor role amounting to only 3.5% on average, with very low variation across regions and time.

<sup>24</sup>This selection excludes non-grape growing regions. For six regions single missing data points until 2003 are imputed using Multivariate Imputation by Chained Equations (MICE, see Buuren & Groothuis-Oudshoorn, 2011). Few extreme observations are also handled with transformation.

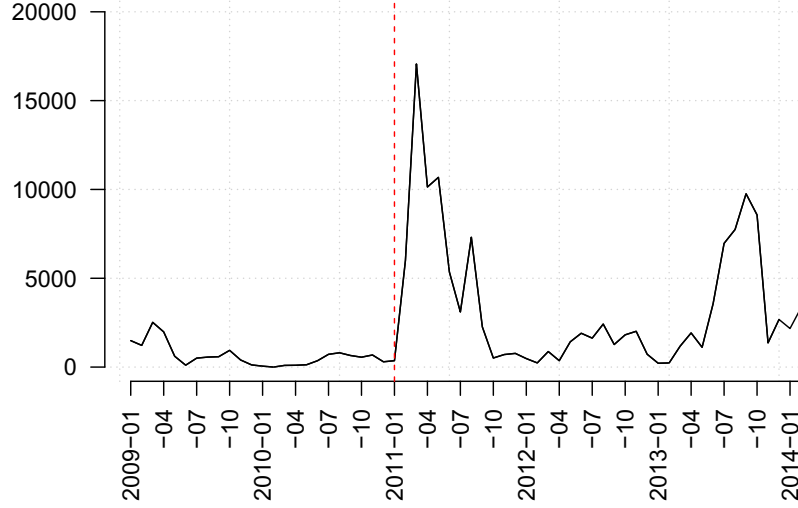


Figure 1: Detected illegal border crossings on the Central Mediterranean route  
(own illustration, source: FRONTEX, 2016)

production (Baccaglio, 2016). Thus, price fluctuations can have only limited impact. Second, no sizable price shocks among Sicilian and Apulian PDOs have been registered (ISMEA, 2018). Third, average PDO wine prices follow similar trends in all regions of Italy and France,<sup>25</sup> decreasing in 2009 due to a lower international demand, and evolving fairly stable until 2012 when they increase due to scarce harvests (FranceAgriMer, 2014; ISMEA, 2018, and see Figure 3 in the Appendix).<sup>26</sup>

Our explanatory variables are chosen to be exogenous, and contain measures of capital, land, labor, as well as other farm and labor market characteristics. Regarding capital, we include capital intensity in terms of book values of machinery over total vineyard hectares (in logs, *lnMachinery*) as well as the lagged investment rate computed as gross investment over total fixed assets (*invRate*). In fact, the degree of capital intensity determines both workers' day-long fatigue level and the efficiency of grower management practices, two main drivers of

<sup>25</sup>The assumption of parallel trends across regions is not rejected (F-test p-value = 0.7).

<sup>26</sup>In fact, in 2012, both France and Italy had similar growing conditions with wet spring and hot summer (WineSpectator, 2012). Extreme cases are 2012 hail storms in Burgundy and Beaujolais that caused particularly low yields and high prices, and Veneto's Prosecco boom between 2009 and 2012.

labor productivity in grape harvesting and grapevine pruning (Lamouria *et al.*, 1963). Further, these variables control for the more intense use of grape harvesting machines in some regions. Land input is included as total vineyard hectares (in logs, *lnLand*) to capture potential returns to scale. To account for different degrees of vineyard specialization at regional level, we use the share of vineyard hectares over total utilized agricultural area (*vineSpec*). Furthermore, we include two variables to control for potential competition on the labor market: the deviation of the regional unemployment rate from its long-term mean (*unempRate*), and the share of population above the age of 15 with less than primary or secondary education (ISCED11) (*unskilledLab*) as an indicator of unskilled workforce available for the harvesting. To account for weather effects that may impact both potential output and harvested output through workers' fatigue, we include the average of the daily minimum temperatures measured at the major regional airports (*minTemp*).<sup>27</sup> Finally, three additional control variables are included: A fixed effect for France to account for unobserved systematic differences in labor productivity at country-level (*france*), a linear time trend (*trend*), and a time dummy to absorb the effects on both vineyard output and hours worked of 2002 anomalous weather events including heavy precipitations that destroyed considerable shares of the harvest in both Italy and France (*weather<sub>2002</sub>*). The treatment dummy,  $D_{it}$ , used to estimate the treatment effect,  $\delta$ , is the interaction term of *treated*  $\times$  *post*, where *treated* is a unit dummy equal to one for Sicily and Apulia, and *post* is a time dummy equal to one for the post-treatment periods.

Table 2 reports data descriptive statistics separately for the treated and control units. In particular, the table shows that variable values for Sicily and Apulia are within the range of the control regions, and mostly around the sample mean. This can be also visualized from the boxplots in the Appendix (Figure 4) displaying the distribution of data. The figure shows overlap of support for individual variable values for Sicily and Apulia, which is especially important when estimating the average treatment effect for the two regions.<sup>28</sup> In particular, the dependent variable for the treated units lies between the first and the second quartile of the sample.

---

<sup>27</sup>This data is scraped from [www.wunderground.com](http://www.wunderground.com).

<sup>28</sup>Only the maximum value of *minTemp* for Sicily exceeds the range slightly.



Table 2: Descriptive Statistics

|                     | Sicily & Apulia |        |        |       | Control units |        |       |       |
|---------------------|-----------------|--------|--------|-------|---------------|--------|-------|-------|
|                     | Mean            | Min    | Max    | SD    | Mean          | Min    | Max   | SD    |
| <i>LabProd</i>      | 2.441           | 2.170  | 2.840  | 0.179 | 3.089         | 1.574  | 4.395 | 0.599 |
| <i>lnLand</i>       | 1.534           | 1.092  | 1.887  | 0.195 | 2.035         | 0.350  | 3.301 | 0.751 |
| <i>vineSpec</i>     | 0.671           | 0.472  | 0.821  | 0.099 | 0.640         | 0.260  | 0.961 | 0.153 |
| <i>lnMachinery</i>  | 8.027           | 7.396  | 8.530  | 0.366 | 8.372         | 6.326  | 9.539 | 0.660 |
| <i>minTemp</i>      | 10.596          | 8.333  | 13.143 | 1.456 | 6.134         | 1      | 12    | 2.552 |
| <i>invRate</i>      | 0.010           | 0.000  | 0.075  | 0.017 | 0.068         | -0.044 | 0.255 | 0.064 |
| <i>unempRate</i>    | 0               | -0.046 | 0.068  | 0.032 | 0             | -0.059 | 0.063 | 0.018 |
| <i>unskilledLab</i> | 0.200           | 0.160  | 0.230  | 0.023 | 0.171         | 0.086  | 0.270 | 0.044 |

## 4 Results

We estimate the causal effects of the 2011 migration wave on vineyard labor productivity using a dynamic regression model as outlined in Section 2. After assessing the validity of our dynamic model specification, we estimate the DPOLS model in equation 1 with and without common time effects. Results from these models show that the 2011 migration wave led to a statistically significant increase in labor productivity of, on average, 11% in 2011 and 2012. Further, robustness checks by the Anderson-Hsiao type regression, and the post-lasso approach confirm the results outlined above.<sup>29</sup>

Table 3 reports the results of the DPOLS model. Models M1, M2 refer to DPOLS without time effects, with time effects (unit-demeaned), respectively. The DPOLS models with/without time effects show that the average causal effect estimate over 2011-2012, i.e., the ATT calculated as  $\exp(\hat{\delta}) - 1$ , ranges between 10.7% and 11.7%, and it is statistically significant at 1%. Considering the average causal estimate from these two models as its lower- and upper-bound, respectively, this means that the 2011 migration wave led to an abnormal average increase in *LabProd* of about 11.2%.<sup>30</sup>

The other explanatory variables show the expected signs. *lnLand* and *lnMachinery* indicate that more land and capital assets increase *LabProd*, pointing

<sup>29</sup>For the analysis, we use the software R, and, particularly, the packages *plm* (Croissant & Millo, 2008), *MatchIt* (Ho *et al.*, 2011), and *hdm* (Chernozhukov *et al.*, 2016).

<sup>30</sup>Static fixed effect and simple pooling regressions estimate relatively larger  $\delta$  of 0.256 and 0.227 with standard errors equal to 0.116 and 0.132 (5% and 10% statistical significance), respectively. This is expected as fixed effects and LDV models bracket the treatment effect between an upper and a lower bound, respectively (Angrist & Pischke, 2010).

Table 3: DPOLS model without time effects (M1), with time effects (M2)

|                             | (M1)                     | (M2)             |
|-----------------------------|--------------------------|------------------|
| $\rho$                      | 0.714*** (0.040)         | 0.723*** (0.041) |
| <i>lnLand</i>               | 0.149*** (0.030)         | 0.136*** (0.029) |
| <i>vineSpec</i>             | 0.018 (0.064)            | -0.011 (0.059)   |
| <i>lnMachinery</i>          | 0.119*** (0.028)         | 0.115*** (0.027) |
| <i>minTemp</i>              | -0.008 (0.005)           | -0.006 (0.005)   |
| <i>invRate</i>              | -0.191 (0.227)           | -0.226 (0.237)   |
| <i>unempRate</i>            | -0.924* (0.546)          | -0.278 (0.663)   |
| <i>unskilledLab</i>         | 1.508** (0.604)          | 1.242** (0.604)  |
| <i>france</i>               | 0.278*** (0.072)         | 0.272*** (0.072) |
| <i>weath<sub>2002</sub></i> | 0.133*** (0.036)         |                  |
| <i>trend</i>                | 0.009** (0.004)          |                  |
| <i>treated</i>              | 0.017 (0.037)            | 0.008 (0.036)    |
| <i>post</i>                 | 0.074** (0.037)          |                  |
| $\delta$                    | 0.102** (0.040)          | 0.111*** (0.038) |
| constant                    | -0.916*** (0.342)        |                  |
| Obs.                        | 350                      | 350              |
| Time effects                | No                       | Yes              |
| F Statistic                 | 450.2***                 | 554.5***         |
| Adjusted R <sup>2</sup>     | 0.918                    | 0.914            |
| Notes:                      | ***p=.01; **p=.05; *p=.1 |                  |

toward economies of scale. On the contrary, larger deviations from the long-term unemployment rate, *unemp*, are negatively related to *LabProd*. Given the systematic higher unemployment rate in the south of both France and Italy, as well as the relatively low variance of such variable over time, this partial effect may simply reflect the north-south gap in labor productivities. A higher availability of unskilled labor positively correlates with *Labprod*, likely indicating the higher labor productivity of more developed regional low-skilled labor markets. Lastly, *france* indicates that *Labprod* is on average higher in France than in Italy. The parameter for the 2002 extreme weather events also has the expected sign and positively affects *Labprod*, mainly due to the drop in labor force needed on vineyards in that year. Lastly, the autoregressive parameter  $\rho$  is statistically significant and amounts to about 0.7, confirming the presence of a well-behaved autoregressive process with a relatively high degree of state dependence.<sup>31</sup>

<sup>31</sup>The key stability conditions of the autoregressive outcome process are initially tested, and are fulfilled, i.e., the outcome covariance structure and unit-root tests indicate weak dependence and covariance stationarity, respectively (Choi, 2001; Hadri, 2000).

To provide evidence supporting the validity of the models and results outlined above, we conduct a series of tests. First, we test a static fixed effect (within) model for the residual presence of unobserved unit-specific effects by means of LM tests (Breusch & Pagan, 1980; Honda, 1985). Results indicate a significant, large residual variance across units (all p-values < 0.01), i.e., the presence of leftover unobserved unit-specific effects also after time-demeaning the model. Moreover, tests show that serial correlation of residuals is present (Breusch, 1978; Godfrey, 1978). Second, we perform the same tests on the dynamic model (equation 1). In this case, the lagged dependent variable absorbs these unobserved unit-specific effects (p-values > 0.1) and residuals' first-order serial correlation. Thus, parameter estimates of the DPOLS model are consistent. Moreover, no cross-sectional dependence is detected (p-values > 0.06, Pesaran, 2004), and Newey-West corrected standard errors (Newey & West, 1994) are used as a safeguard also with rejected serial correlation (as suggested by Wooldridge, 2013, ch. 12.5).

We also test our results for robustness against different choices of the control group. Based on covariate balance, we select the  $k$  best control matches (nearest neighbors) for each treated region, and we re-estimate the DiD model for different values of  $k$ . We find that the DiD parameter is generally of similar magnitude and statistically significant (see Table 8 in the Appendix).<sup>32</sup>

Next, we transform our estimates on labor productivity into estimates of the unreported (illegal) hours worked at Sicily's and Apulia's vineyards. To do so, we perform simple back of the envelope calculations in the following way. Total production *Output* is a function of labor productivity *LabProd* and labor input  $L$  such that  $Output = LabProd^l * L^l + LabProd^{il} * L^{il}$  with superscripts  $l$  and  $il$  denoting legal and illegal labor input, respectively. Solving for  $L^{il}$  delivers the illegal input as a function of the observed values of output and legal labor input, while the values of true labor productivity of legal and illegal input are not observed. An estimate of the true labor productivity of legal input is calculated from the reported labor productivity  $LP^{reported}$  using the estimated average treatment effect:  $Lab\hat{Prod}^l = e^{-\hat{\delta}} * LabProd^{reported}$ . Thereby, we assume homogenous treatment effects for the treated regions. Further, we parameterize labor productivity of illegal input as a function of legal inputs as

---

<sup>32</sup>Full results for in-time placebos, further dynamic and static models are available upon request.

$LabProd^{il} = \theta LabProd^l$ . This delivers:

$$\hat{L}^{il} = \frac{L^l \times (LabProd^{reported} - Lab\hat{P}rod^l)}{\theta Lab\hat{P}rod^l} = \frac{L^l (LabProd^{reported} [1 - e^{-\hat{\delta}}])}{\theta e^{-\hat{\delta}} * LabProd^{reported}} \quad (2)$$

The rationale behind this parameterization is that the productivity of illegal labor might differ from that of legal labor, e.g., due to fatigue from long working hours (Palmisano & Sagnet, 2016; Lamouria *et al.*, 1963). Therefore,  $\theta$  relates the two labor productivities such that illegal labor and legal labor are identically productive for  $\theta = 1$ , and for example  $\theta = 1.2$  (0.8) indicates 20% higher (lower) labor productivity for illegal labor. Further, not only is the relationship between  $\hat{L}^{il}$  and  $\theta$  non-linear, but  $\hat{L}^{il}$  decreases faster for low productivity levels than for higher levels ( $\frac{\partial^2 \hat{L}^{il}}{\partial \theta^2} > 0$ ).

Table 4 reports the estimates of unreported work hours for the average farm, the two different model specifications, and for different values of  $\theta$  arbitrarily chosen to vary between 0.8 and 1.2. Estimates are calculated under the assumption of homogeneous treatment effects, i.e.,  $\delta$  is constant across the treated units. Estimates slightly vary between years and regions. We find generally slightly higher values for Apulia than for Sicily, although differences are small. For identical productivity of legal and illegal workforce ( $\theta = 1$ ), unreported labor input is estimated to vary between 205 and 278 hours. These estimates correspond to 11% (205h/1800h) and 15% (278h/1800h) of an annual work unit (AWU) – which is the agricultural equivalent of a full-time employee as defined by the EU (EC, 2017) – or one person working around 5 to 7 40-hour weeks.<sup>33</sup> This corresponds to the length of the harvest for one grape variety (about 30 days), thus about one full-time grape picker might not be reported by the average farm during this time.

While these estimates seem to be rather low, it should be noted that the sector is characterized by a strong fragmentation with a large number of fairly small vineyards. To estimate the overall effect, we transform our estimates at a regional level by multiplication with the number of farms within a region. Table 4 summarizes the results. For  $\theta = 1$ , total unreported labor input ranges between 9 and 10.3 million hours in 2011, and 10 and 12.2 million hours in

---

<sup>33</sup>However, for illegal labor input, and in particular for labor input employed through caporalato, working conditions are usually tougher, and 40-hour weeks are likely an underestimate (INEA, 2012).

2012. Annual estimates range between 12.2 and 13.6 million hours for  $\theta = 0.8$ , and between 8 and 9 million hours for  $\theta = 1.2$ . In terms of AWU (assumed to work 1800 hours a year), 10 million unreported hours correspond to around 5,500 illegally employed AWUs, or 5,500 full-time employees.

However, this calculation is sensitive to various factors: First, hours worked per day might be considerably higher. Second, estimates vary strongly with  $\theta$ . Thirdly, the estimated average treatment effect,  $\hat{\delta}$ , identifies only the increase in labor productivity due the migration wave. If labor productivity is already overestimated before this shock due to unreported labor, the estimated unreported hours are only a lower bound of the actual numbers, which is likely the case due to the historical presence of illegal employment of labor (Flai-Cgil, 2016).

Table 4: Estimated unreported hours worked for the average farm by region (and total per region in thousand)

|                                      | $\theta = 0.8$ | $\theta = 1$ | $\theta = 1.2$ |
|--------------------------------------|----------------|--------------|----------------|
| <i>M1</i> ( $\hat{\delta} = 0.102$ ) |                |              |                |
| Sicily 2011                          | 268 (5,825)    | 214 (4,660)  | 178 (3,884)    |
| Apulia 2011                          | 286 (6,064)    | 229 (4,851)  | 190 (4,043)    |
| Sicily 2012                          | 249 (6,490)    | 199 (5,192)  | 166 (4,327)    |
| Apulia 2012                          | 285 (6,061)    | 228 (4,848)  | 190 (4,040)    |
| <i>M2</i> ( $\hat{\delta} = 0.111$ ) |                |              |                |
| Sicily 2011                          | 293 (6,368)    | 234 (5,095)  | 195 (4,246)    |
| Apulia 2011                          | 312 (6,629)    | 250 (5,304)  | 208 (4,420)    |
| Sicily 2012                          | 272 (7,095)    | 217 (5,676)  | 181 (4,730)    |
| Apulia 2012                          | 312 (6,626)    | 250 (5,300)  | 208 (4,417)    |

## 4.1 Robustness checks

### 4.1.1 In-time placebo DiD

Since treatment effects on the outcomes of the treated regions before the migrant wave are not expected, thus, should not occur, we introduce placebo treatment dummies for the five years before the shock. The left graph of Figure 2 displays the evolution of the placebo treatment effects over time (x-axis), and the respective parameter estimates as well as confidence intervals (y-axis). For 2011 and 2012, the estimated average causal effects result to be the largest in magnitude and with the smallest confidence levels. Joint (as well as indi-

vidual) statistical insignificance is not rejected with a  $p\text{-value} = 0.99$ . Instead, yearly treatment effects for 2011 and 2012 are statistically significant at 5%. On the right, instead, parameter estimates (x-axis) are plotted against their respective standard errors (y-axis), and the shaded area indicates that there are no causal effects more extreme than those for 2011 and 2012.

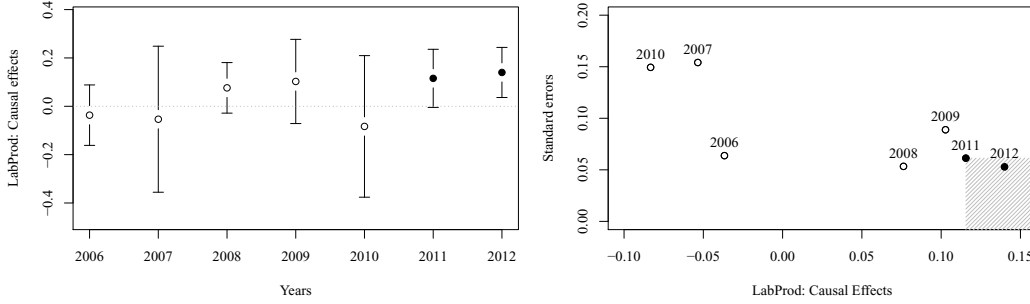


Figure 2: DiD in-time placebos and dynamic causal effects

#### 4.1.2 Anderson-Hsiao type regression

To further check for the robustness of our results against unobserved fixed effects,  $\mu_i$ , we use the AH estimator on the first-differenced dynamic model (Anderson & Hsiao, 1981). First differencing swaps away unit-specific fixed effects; however, it induces correlation between the first-differenced LDV and error term, thus, endogeneity. We choose the second and third lagged outcomes to instrument the first-differenced lagged outcome, both in levels and in first differences. Table 5 reports the results of the two-stage least square estimation using first-differenced instruments (column 1 and 2) and level instruments (column 3 and 4), respectively, with and without time effects. All four models show a positive causal effect of the illegal labor supply shock on measured labor productivity. Parameter estimates of these causal effects are generally of higher magnitude than those obtained with the DPOLS specification (cp. Table 3), ranging between 15.4 and 17.9%. Despite the overall lower efficiency of the AH estimator, these estimates are statistically significant at the 1% level. The instruments are relevant according to robust F-tests on the first-stage regressions ( $p\text{-values} < 0.01$ ). Further, auxiliary-regression based Sargan tests with heteroskedasticity and autocorrelation corrected residuals indicate that the instruments are exogenous, i.e., uncorrelated with the AH model's resid-

uals (p-value > 0.3). However, it should be noted that first differencing the model induces serially correlated residuals, as well as unobserved unit-specific heterogeneity with a large residual variance across regions (p-values < 0.005).<sup>34</sup> For this reason, both the exogeneity of the first-differenced lagged instruments and the model specification may be problematic. Therefore, we consider robustness of AH model estimates only as an indicator of robustness against alternative model specifications.

Table 5: AH model estimates with different IV strategies and with/without time effects

|                      | First-difference IVs     |         |          |         |
|----------------------|--------------------------|---------|----------|---------|
|                      | (AH1)                    |         | (AH2)    |         |
| $\rho$               | 0.217                    | (0.189) | 0.034    | (0.132) |
| $\ln Land$           | 0.535***                 | (0.039) | 0.517*** | (0.048) |
| $\ln vineSpec$       | 0.116                    | (0.241) | 0.003    | (0.275) |
| $\ln Machinery$      | 0.065*                   | (0.039) | 0.064*   | (0.036) |
| $\ln minTemp$        | 0.004                    | (0.010) | 0.013**  | (0.006) |
| $\ln invRate$        | 0.249                    | (0.223) | 0.200    | (0.201) |
| $\ln unempRate$      | -0.336                   | (1.297) | -0.300   | (2.041) |
| $\ln unskilledLab$   | -0.008                   | (0.845) | 0.116    | (0.995) |
| $\ln post$           | 0.026*                   | (0.014) |          |         |
| $\ln weather_{2002}$ | 0.134***                 | (0.017) |          |         |
| $\delta$             | 0.151***                 | (0.039) | 0.143*** | (0.029) |
| Obs.                 | 325                      |         | 325      |         |
| Time effects         | No                       |         | Yes      |         |
| F Statistic          | 9.199***                 |         | 5.394*** |         |
| Adjusted R2          | 0.151                    |         | 0.190    |         |
| Notes:               | ***p=.01; **p=.05; *p=.1 |         |          |         |

<sup>34</sup>In this case, we follow Driscoll & Kraay (1994) and use standard errors corrected for both potential autocorrelation and cross-sectional dependence in the residuals.

| x                    |          |         |          |         |
|----------------------|----------|---------|----------|---------|
| Level IVs            |          |         |          |         |
|                      | (AH3)    |         | (AH4)    |         |
| $\rho$               | 0.317    | (0.270) | 0.146    | (0.154) |
| $\ln Land$           | 0.533*** | (0.038) | 0.517*** | (0.046) |
| $\ln vineSpec$       | 0.138    | (0.246) | 0.017    | (0.300) |
| $\ln Machinery$      | 0.057    | (0.043) | 0.059    | (0.037) |
| $\ln minTemp$        | 0.003    | (0.011) | 0.013**  | (0.007) |
| $\ln invRate$ 0.217  | (0.242)  | 0.171   | (0.218)  |         |
| $\ln unempRate$      | -0.319   | (1.295) | -0.257   | (2.117) |
| $\ln unskilledLab$   | 0.048    | (0.915) | 0.120    | (1.075) |
| $\ln post$           | 0.025*   | (0.015) |          |         |
| $\ln weather_{2002}$ | 0.127*** | (0.023) |          |         |
| $\delta$             | 0.165*** | (0.045) | 0.161*** | (0.032) |
| Obs.                 | 325      |         | 325      |         |
| Time effects         | No       |         | Yes      |         |
| F Statistic          | 8.346*** |         | 4.882*** |         |
| Adjusted R2          | 0.122    |         | 0.099    |         |

Notes: \*\*\*p=.01; \*\*p=.05; \*p=.1

#### 4.1.3 Post-lasso approach

To check for possible model misspecification and to allow for a more flexible functional form, we enlarge the set of covariates with additional farm-related characteristics, as well as high-dimensional variables, in particular, interaction terms, log specifications, and second and third order orthogonal polynomials. We perform, first, model selection by lasso and, then, post-selection DiD estimation of the ATT as proposed by Belloni *et al.* (2012, 2013). Main results from the post-lasso regression are the following: (I) The model specification is dynamic ( $\rho$  is not shrunk to zero); (II) rigorous lasso selects 17 variables, including some of our original covariates, some of their interaction terms, and some newly included farm-related variables; (III) after model selection, post-lasso DPOLS estimation reports plausible signs, and additionally selects the rent paid for farm land and buildings and rental charges (in logs); and (IV) the treatment dummy is highly statistically significant (p-value = 0.015), amounting to 9.3% (for details, see Table 9 in the Appendix).



Table 6 presents a comparison of the ATT estimates obtained by the previous methods, and the post-lasso approach. Although smaller in magnitude, the ATT computed by post-lasso confirms the presence of a statistically significant average causal effect of the illegal labor supply shock on vineyard labor productivity in Sicily and Apulia, indicating the presence of illegal labor employment.

Table 6: Comparison of ATT estimates

| Model             | Estimate | Std. Error |
|-------------------|----------|------------|
| DPOLS: M1         | 0.102    | 0.040      |
| DPOLS: M2         | 0.111    | 0.038      |
| AH2 (lower bound) | 0.143    | 0.029      |
| AH3 (upper bound) | 0.165    | 0.045      |
| Post-lasso        | 0.093    | 0.037      |

## 5 Conclusions

This paper aims to identify and quantify illegal employment causal to an irregular migrant wave, and shows (I) a new way to identify unreported labor, and (II) a lower-bound estimate of the amount of unreported labor caused by the migration-induced supply shock. To identify changes in illegal employment we use the exogenous variation of the 2011 Arab Spring migrant landings on southern Italian shores, and we consider labor productivity on vineyards in the landing regions. Identification relies on abnormal increases in reported labor productivity coming from underreported labor hours. Based on farm-level data aggregated at regional level and using a dynamic panel model, labor productivity is estimated to abnormally increase by about 11% on average for 2011 and 2012 on vineyards in the landing regions. We show that this effect corresponds to around 10 million hours irregularly worked in the treated regions in each year – or around 5,500 agricultural work units. These workers can be undocumented migrants from the 2011 migration wave, and/or other workers available to work illegally, including former legally employed workers. Thereby, our results suggest that illegal workforce displaced legal workforce, leading to underreported labor input and overreported labor productivity.

These results are in line with the literature which finds low-skilled jobs (see, e.g., Dustmann *et al.*, 2016; Peri, 2016) and informal native employment (see, e.g., Tumen, 2016; Del Carpio & Wagner, 2015) the most vulnerable to

migrant labor supply shocks. Indeed, this is the case for vineyard labor: The seasonal nature and the low skill requirements of field picker jobs limit workers' bargaining power and makes them substitutable.

Our results also underline several flaws in existing policies. First, the lack of regulation and inspection on farmlands prevents the effectiveness of existing laws against caporalato and illegal employment because employers face very low probabilities of being caught. Second, in this respect, European laws that assign temporary residence permits to irregular migrants who denounce severe exploitation - partially introduced in Italy only in 2012 - could be fully applied to guarantee workers' protection, and limit employers' exploitation incentives. Third, more efficient evaluations of asylum requests would avoid lengthy, complicated, and often unclear procedures that encourage migrants to stay irregular. Finally, future research should analyze alternatives to the existing voucher systems and other type of contracts designed by European governments that aim to facilitate matching agricultural labor demand and supply.

Also, several questions related to our study remain open and should be addressed in future research. Generally, the impacts of the Arab Spring migration crisis on European labor markets needs further investigation. In particular, in addition to employment effects, the impact of the supply shock on wages of both legal and illegal labor should be analyzed. Further, the current analysis should be extended to the whole agribusiness as it is the sector that absorbs most of the illegal workforce. Finally, long-term effects on labor markets need to be evaluated taking into account the current EU immigration policy and the recent regulatory efforts against labor exploitation.

## Acknowledgements

We thank the American Association of Wine Economists for financial support, and Mariusz Stefan Migas (European Commission), Mauro Santangelo (CREA), and Tiziana Sarnari (ISMEA) for providing data. The paper greatly benefited from discussions at the 2017 conferences of the American Association of Wine Economists (AAWE, Padua), the Agricultural Economics Society (AES, Dublin), the European Workshop on Efficiency and Productivity Analysis (EWEPA, London), and the Econometrics of Panel Data and Net-

work Analysis Summer School (CRC, Berlin), as well as at the seminars of the Loughborough University, University of Verona, and DIW Berlin. We also thank the participants of the Fifth CReAM workshop for suggestions and discussion. Also, we thank thank Pio Baake, Boris Bravo-Ureta, Tomaso Duso, Bernd Fitzenberger, Cristina Salvioni, Jeffrey Wooldridge, and Angelo Zago for fruitful input.

## 6 Appendix

Table 7: Share of asylum seekers hosted in SPRAR reception centers in each Italian region over 2010-2012 (SPRAR, 2010, 2011, 2012)

|                   | 2010 | 2011 | 2012 |                   | 2010 | 2011 | 2012 |
|-------------------|------|------|------|-------------------|------|------|------|
| <b>North-East</b> |      |      |      | <b>North-West</b> |      |      |      |
| Friuli V.G.       | 4.8  | 4.6  | 4.2  | Aosta V.          | 0.0  | 0.0  | 0.0  |
| Veneto            | 4.7  | 5.8  | 4.0  | Piedmont          | 4.6  | 5.3  | 4.5  |
| Trentino A.A.     | 0.6  | 0.6  | 0.4  | Lombardy          | 16.5 | 5.7  | 16.8 |
| Emilia R.         | 6.2  | 7.8  | 6.8  | Liguria           | 2.7  | 3.0  | 2.2  |
| <b>Center</b>     |      |      |      | <b>South</b>      |      |      |      |
| Tuscany           | 4.4  | 4.5  | 4.6  | Abruzzo           | 0.5  | 0.5  | 0.6  |
| Marche            | 4.2  | 4.5  | 3.5  | Molise            | 0.5  | 0.6  | 0.6  |
| Lazio             | 22.4 | 26.2 | 21.2 | Campania          | 2.9  | 3.2  | 2.0  |
| Umbria            | 2.0  | 2.7  | 2.0  | Basilicata        | 0.7  | 0.6  | 0.7  |
| <b>Islands</b>    |      |      |      | Apulia            | 7.1  | 8.0  | 6.2  |
| Sicily            | 11.4 | 11.3 | 14.6 | Calabria          | 3.5  | 4.7  | 4.9  |
| Sardinia          | 0.4  | 0.4  | 0.3  | <b>Total</b>      | 7056 | 7598 | 7823 |

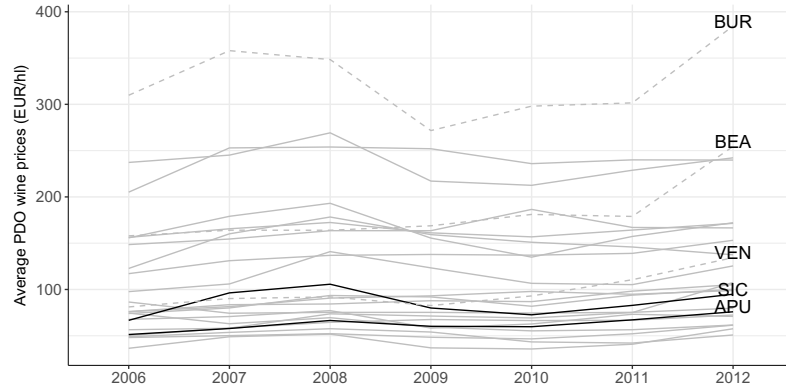


Figure 3: Average high-quality (PDO) wine prices for the treated regions Sicily (SIC) and Apulia (APU) (black) and control regions (grey), highlighting the discussed cases of Burgundy (BUR), Beaujolais (BEA), and Veneto (VEN) (dashed) (own illustration, source: ISMEA, 2018; FranceAgriMer, 2014).

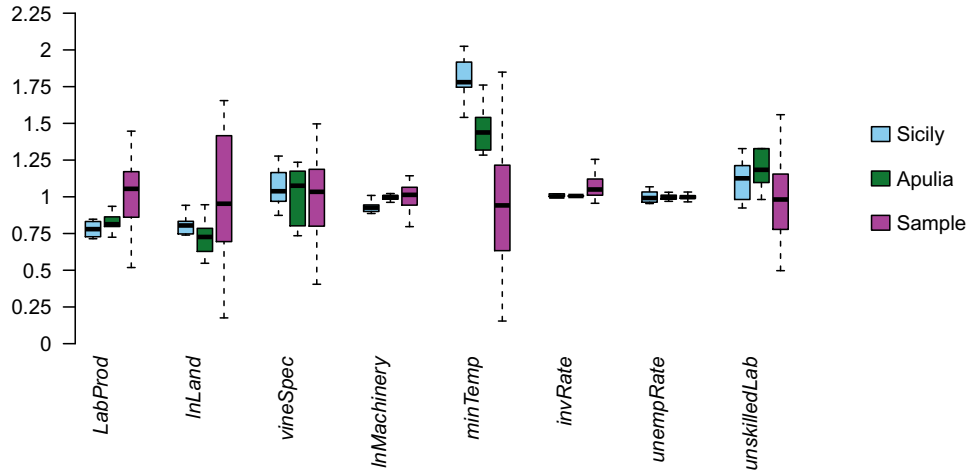


Figure 4: Descriptive statistics: Boxplots of mean-corrected variables (for comparability and visualization reasons, each variable is normalized by dividing it by its mean)

Table 8: DPOLS model estimates with time fixed effects (M2) for different choices of the control group based on  $k$  nearest neighbor matching (1999-2012). For each treated region,  $k$  nearest neighbors are selected resulting in sample sizes of  $2(k+1)T$ , and the full control group is selected for  $k = 12$ .

|                         | ( $k = 6$ )              | ( $k = 7$ )         | ( $k = 8$ )         | ( $k = 9$ )         | ( $k = 10$ )        | ( $k = 11$ )        |
|-------------------------|--------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| $\rho$                  | 0.665***<br>(0.059)      | 0.634***<br>(0.066) | 0.641***<br>(0.050) | 0.682***<br>(0.051) | 0.673***<br>(0.043) | 0.719***<br>(0.042) |
| $\ln Land$              | 0.121**<br>(0.053)       | 0.170***<br>(0.049) | 0.187***<br>(0.036) | 0.202***<br>(0.040) | 0.203***<br>(0.030) | 0.141***<br>(0.029) |
| $\text{vineSpec}$       | -0.006<br>(0.137)        | -0.097<br>(0.084)   | 0.040<br>(0.077)    | 0.046<br>(0.068)    | 0.096<br>(0.068)    | 0.009<br>(0.059)    |
| $\ln Machinery$         | 0.135***<br>(0.036)      | 0.158***<br>(0.040) | 0.106***<br>(0.035) | 0.116***<br>(0.029) | 0.112***<br>(0.027) | 0.119***<br>(0.027) |
| $\text{minTemp}$        | -0.008<br>(0.008)        | -0.006<br>(0.006)   | -0.015**<br>(0.007) | -0.009<br>(0.005)   | -0.011**<br>(0.006) | -0.007<br>(0.005)   |
| $\text{invRate}$        | -0.414<br>(0.427)        | -0.273<br>(0.278)   | 0.267<br>(0.352)    | 0.157<br>(0.286)    | 0.196<br>(0.301)    | -0.233<br>(0.244)   |
| $\text{unempRate}$      | -0.059<br>(0.852)        | 0.420<br>(0.735)    | -0.250<br>(0.729)   | -0.326<br>(0.691)   | -0.523<br>(0.693)   | -0.364<br>(0.669)   |
| $\text{unskilledLab}$   | 0.310<br>(0.943)         | -0.058<br>(0.677)   | 0.045<br>(0.762)    | 1.197*<br>(0.666)   | 1.320**<br>(0.653)  | 1.394**<br>(0.607)  |
| $\text{france}$         | 0.274***<br>(0.102)      | 0.277***<br>(0.083) | 0.072<br>(0.089)    | 0.154*<br>(0.080)   | 0.158**<br>(0.077)  | 0.277***<br>(0.074) |
| $\text{treated}$        | -0.024<br>(0.042)        | 0.004<br>(0.038)    | -0.018<br>(0.040)   | 0.008<br>(0.035)    | 0.010<br>(0.036)    | 0.011<br>(0.036)    |
| $\delta$                | 0.135**<br>(0.053)       | 0.105**<br>(0.051)  | 0.115***<br>(0.043) | 0.093**<br>(0.040)  | 0.104***<br>(0.039) | 0.118***<br>(0.038) |
| Obs.                    | 196                      | 224                 | 252                 | 280                 | 308                 | 336                 |
| F Statistic             | 175.8***                 | 272.4***            | 347.2***            | 358.7***            | 504.6***            | 545.6***            |
| Adjusted R <sup>2</sup> | 0.905                    | 0.892               | 0.912               | 0.897               | 0.906               | 0.915               |
| Notes:                  | ***p=.01; **p=.05; *p=.1 |                     |                     |                     |                     |                     |

Table 9: Post-lasso model estimates

|                         | Post-lasso               | Std. Error |
|-------------------------|--------------------------|------------|
| $\rho$                  | 0.630***                 | (0.049)    |
| $\ln Land * \rho$       | -0.004                   | (0.007)    |
| $vineSpec * \rho$       | -0.014                   | (0.022)    |
| $invRate * unempRate$   | -10.115*                 | (5.235)    |
| $weather_{2002}$        | 0.109***                 | (0.035)    |
| $france$                | -0.024                   | (0.063)    |
| $dummy_{2000}$          | -0.073**                 | (0.031)    |
| $dummy_{2012}$          | 0.041                    | (0.030)    |
| $trend$                 | 0.003                    | (0.004)    |
| $\ln Rent$              | 0.040**                  | (0.019)    |
| $\ln Depreciation$      | 0.057                    | (0.049)    |
| $\ln Taxes$             | 0.050*                   | (0.026)    |
| $\ln OtherInputs$       | 0.011                    | (0.028)    |
| $\ln MachineryCosts$    | 0.052*                   | (0.027)    |
| $otherInputs^2$         | 0.053                    | (0.125)    |
| $otherInputs^3$         | -0.110                   | (0.130)    |
| $post$                  | 0.012                    | (0.037)    |
| $treated$               | 0.024                    | (0.037)    |
| $\delta$                | 0.089**                  | (0.037)    |
| constant                | -0.552**                 | (0.273)    |
| Obs.                    | 350                      |            |
| F Statistic             | 482.8***                 |            |
| Adjusted R <sup>2</sup> | 0.922                    |            |
| Notes:                  | ***p=.01; **p=.05; *p=.1 |            |

## References

- Achen, C. 2000. Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables. *Prepared for the Annual Meeting of the Political Methodology Section of the American Political Science Association, UCLA, July 20-22.*
- ANCI. 2011. Accoglienza ai profughi provenienti dal Nord Africa. *National Association of Italian Municipalities - Anci. Available at [http://www.anci.lombardia.it/documenti/ACCOGLIENZA\\_AI\\_PROFUGHI\\_PROVENIENTI\\_DAL\\_NORD\\_AFRICA\\_30\\_maggio.doc](http://www.anci.lombardia.it/documenti/ACCOGLIENZA_AI_PROFUGHI_PROVENIENTI_DAL_NORD_AFRICA_30_maggio.doc).*
- Anderson, T. W., & Hsiao, Cheng. 1981. Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association*, **76**(375), 598–606.
- Angrist, J. D., & Pischke, J.-S. 2010. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- ASGI. 2011. Grave preoccupazione per le ripetute violazioni del diritto nei riguardi degli stranieri respinti, espulsi o trattenuti nei CIE, dei richiedenti asilo e dei lavoratori stranieri

- (12 August 2011). *Association for Juridical Studies on Immigration (ASGI)*, Retrieved from <http://old.asgi.it/>.
- ASGI. 2012. Il diritto alla protezione. Studio sullo stato del sistema di asilo in Italia e proposte per una sua evoluzione. *European Refugee Fund 2008-2013, Association for Juridical Studies on Immigration (ASGI)*.
- ASGI. 2015. Sanctions against employers of illegally staying third-country nationals. *Association for Juridical Studies on Immigration (ASGI)*, Retrieved from <http://old.asgi.it/>.
- Assosomm. 2016. Attiviamo Lavoro. Le potenzialità del lavoro in somministrazione per il settore dell'agricoltura. *Report of The European House-Ambrosetti for the Italian Association of Labor Agencies (Assosomm)*.
- Autor, D. 2003. Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics*, **21**(1), 1–42.
- Baccaglio, M. 2016. Produzione di vino e superfici vitate - aggiornamento 2011. Available at <http://www.inumeridelvino.it/2012/07/sicilia-produzione-di-vino-e-superfici-vitate-aggiornamento-2011.html>.
- Balkan, B., & Tumen, S. 2016. Immigration and prices: Quasi-experimental evidence from Syrian refugees in Turkey. *Journal of Population Economics*, **29**(3), 657–686.
- Belloni, A., & Chernozhukov, V. 2013. Least squares after model selection in high-dimensional sparse models. *Bernoulli*, **19**(2), 521–547.
- Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, **80**(6), 2369–2429.
- Belloni, A., Chernozhukov, V., & Hansen, C. 2013. Inference for High-Dimensional Sparse Econometric Models. *Advances in Economics and Econometrics, Tenth World Congress, Vol. 3: Econometrics, Cambridge University Press: Cambridge*, 245–295.
- Borjas, George. 2017. The labor supply of undocumented immigrants. *Labour Economics*, **46**, 1–13.
- Boswell, Christina, & Straubhaar, Thomas. 2004. The Illegal Employment of Foreign Workers: an Overview. *Intereconomics*, **39**(1), 4–7.
- Breusch, T. S. 1978. TESTING FOR AUTOCORRELATION IN DYNAMIC LINEAR MODELS. *Australian Economic Papers*, **17**(31), 334–355.
- Breusch, T.S., & Pagan, A.R. 1980. The Lagrange Multiplier Test and its applications to model specification in econometrics. *Review of Economic Studies*, **47**(1), 239–253.

- Buuren, Stef, & Groothuis-Oudshoorn, Karin. 2011. mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, **45**(3).
- Ceritoglu, Evren, Yunculer, Gurcihan, Burcu, H, Torun, Huzeyfe, & Tumen, Semih. 2015. The impact of Syrian refugees on natives' labor market outcomes in Turkey: Evidence from a quasi-experimental design. *IZA Discussion Paper*, **9348**.
- CeSPI. 2012. L'impatto delle primavere arabe sui flussi migratori regionali e verso l'Italia. *Italian Observatory of International Politics, CeSPI, Report n. 59*. Retrieved from <http://www.parlamento.it>.
- Chernozhukov, V., Hansen, C., & Spindler, M. 2016. hdm: High-Dimensional Metrics. *R Journal*, **8**(2), 185–199.
- Chernozhukov, V., Goldman, M., Semenova, V., & Taddy, M. 2017. Orthogonal Machine Learning for Demand Estimation: High Dimensional Causal Inference in Dynamic Panels. *ArXiv e-prints (arXiv:1712.09988)*. Available at <http://adsabs.harvard.edu/abs/2017arXiv171209988C>.
- Chernozhukov, Victor, Chetverikov, Denis, & Kato, Kengo. 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, **41**(6), 2786–2819.
- Choi, I. 2001. Unit root tests for panel data. *Journal of International Money and Finance*, **20**(2), 249–272.
- CMO. 2008. Regulation for Common Market Organization 479/2008. *European Common Organization of Agricultural Markets (CMO)*.
- Croissant, Y., & Millo, G. 2008. Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, **27**(2), 1–43.
- Del Carpio, Ximena V, & Wagner, Mathis C. 2015. The impact of Syrian refugees on the Turkish labor market. *World Bank Policy Research Working Paper*, **7402**.
- DPC. 2017. Piano per l'accoglienza dei migranti (12 April 2011). *Department of Civil Protection (DPC) of the Presidency of the Council of Ministers*.
- Driscoll, J. C., & Kraay, A. C. 1994. Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, **80**(4), 549–560.
- Dustmann, Christian, Schönberg, Uta, & Stuhler, Jan. 2016. The Impact of Immigration: Why Do Studies Reach Such Different Results? *Journal of Economic Perspectives*, **30**(4), 31–56.
- EC. 2017. Farm Accountancy Data Network (European Commission). Available at <http://ec.europa.eu/agriculture/rica/index.cfm>.



- ETI. 2015. Decent work in Italian agriculture: Counteracting exploitation of migrant workers in tomato production. *Project of Ethical Trading Initiative Norway (IEH), Ethical Trading Initiative (ETI) and Danish Ethical Trading Initiative (DIEH)*.
- EUROSTAT. 2017. *Labor Force Survey*. Tech. rept. Eurostat.
- Fitzenberger, B., Bergemann, A., & Speckesser, S. 2009. Evaluating the dynamic employment effects of training programs in East Germany using conditional differences-in-differences. *Journal of Applied Econometrics*, **24**(5), 797–823.
- Flai-Cgil. 2012. Agromafias and Caporalato, First Report. *Osservatorio Placido Rizzotto Flai-Cgil*.
- Flai-Cgil. 2014. Agromafias and Caporalato, Second Report. *Osservatorio Placido Rizzotto Flai-Cgil*.
- Flai-Cgil. 2016. Agromafias and Caporalato, Third Report. *Osservatorio Placido Rizzotto*.
- Flai-Cgil. 2018. Agromafias and Caporalato, Fourth Report. *Osservatorio Placido Rizzotto Flai-Cgil*.
- FranceAgriMer. 2014. Les chiffres de la filière viti-vinicole 2002/2012. *The National Authority for Agriculture and Sea Products, FranceAgriMer*.
- FRONTEX. 2016. Migratory Routes Map. *European Border and Coast Guard Agency*. Available at <http://frontex.europa.eu/along-eu-borders/migratory-map/>.
- Gaeta, D., & Corsinovi, P. 2014. *Economics, Governance, and Politics in the Wine Market: European Union Developments*. Palgrave Macmillan US.
- Gatti, Fabrizio. 2015. Migranti, in centomila sono scomparsi (21 January 2015). *L'Espresso*. Available at <http://espresso.repubblica.it/plus/articoli/2015/01/21/news/migranti-la-grande-fuga-1.195858>.
- Giangrande, A. 2017. *Caporalato Ipocrisia e Speculazione*. Vol. 165. L'Italia del Trucco, l'Italia che siamo.
- Godfrey, L. G. 1978. Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*, **46**(6), 1293–1301.
- Gola, Eleonora Ghizzi. 2015. L'accoglienza dei richiedenti e titolari di protezione internazionale in Italia. Aspetti giuridici e sociologici. *La Rivista, ADIR Research Center, University of Firenze*, Retrieved from <http://www.adir.unifi.it/rivista>.
- Hadri, K. 2000. Testing for Stationarity in Heterogeneous Panel Data. *Econometrics Journal*, **3**(2), 148–161.

- Heckman, James. 1981. *Heterogeneity and State Dependence*. National Bureau of Economic Research.
- Ho, Daniel E., Imai, Kosuke, King, Gary, & Stuart, Elizabeth A. 2011. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, **42**(8), 1–28.
- Honda, J. 1985. Testing the Error Components Model with Non-normal Disturbances. *Review of Economics Studies*, **52**, 681–690.
- Hsiao, Cheng. 2014. *Dynamic Models with Variable Intercepts*. 3 edn. Econometric Society Monographs. Cambridge University Press.
- Il Post. 2011. Il mini-accordo tra Italia e Tunisia (6 April 2011). Retrieved from <http://www.ilpost.it/>.
- INEA. 2012. Indagine sull’impiego degli immigrati in agricoltura in Italia. *Report of the National Institute of Agricultural Economics, INEA*.
- INEA. 2014. Indagine sull’impiego degli immigrati in agricoltura in Italia. *Report of the National Institute of Agricultural Economics, INEA*.
- Interior Ministry. 2012. Emergenza Nord Africa - Procedura informatizzata Vestanet C3 - gestione Nord Africa. *Act of 30 October 2012, n. 5426*.
- ISMEA. 2018. Data on Italian PDO wine prices between 2006 and 2012 [obtained upon request]. *The Italian Institute for Services to the Agricultural and Food Markets, ISMEA*.
- ISMU. 2015. Sbarchi. Serie storica anni 2011-2014. *Institut for Multiethnic Studies (ISMU). Data available at [http://ismu.org/wp-content/uploads/2015/01/Sbarchi\\_serie-2011-2014.xls](http://ismu.org/wp-content/uploads/2015/01/Sbarchi_serie-2011-2014.xls)*.
- ISTAT. 2017. Economia non osservata nei conti nazionali. *ISTAT National Statistics*.
- Jorgenson, D. W., & Griliches, Z. 1967. The Explanation of Productivity Change. *Review of Economic Studies*, **34**(3), 249–283.
- Keele, Luke, & Kelly, Nathan J. 2006. Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables. *Political Analysis*, **14**(2), 186–205.
- Kelly, Charles B. 1977. Counting the Uncountable: Estimates of Undocumented Aliens in the United States. *Population and Development Review*, **3**(4), 473–481.
- Labanca, Claudio. 2016. The effects of a temporary migration shock: Evidence from the Arab Spring migration towards Italy. *University of California at San Diego, Economics Working Paper Series*.

- Lambruschi, Paolo. 2012. Integrazione alla prova. La Caritas: Non rimandare all'inferno chi chiede asilo (20 June 2012). *Avvenire*.
- Lamouria, L.H., Studer, H.E., & Brewer, H.L. 1963. Improving the productivity of pruning labor in the vineyard. *California Agriculture*, **17**(3), 2–3.
- Lianos, Theodore P., Sarris, Alexander H., & Katseli, Louka T. 1996. Illegal Immigration and Local Labour Markets: The Case of Northern Greece. *International Migration*, **34**(3), 449–484.
- Meyer, Antoine, & Dumortier, Thomas. 2014. Severe forms of Labour Exploitation: Supporting victims of severe forms of labour exploitation in having access to justice in EU Member States. *European Union Agency for Fundamental Rights (FRANET) and French Institute on Rights and Freedoms (IFDL)*.
- MPP. 2012. Emergenza Nordafrica - Prorogati i permessi umanitari ai tunisini. E i 25.000 richiedenti asilo provenienti dalla Libia? (22 May 2012). *Melting Pot Project, Law Archive*. Retrieved from <http://www.meltingpot.org>.
- MPP. 2013. Dalla protezione temporanea alla protezione internazionale: Quale accoglienza? *Melting Pot Project*. Available at [http://www.meltingpot.org/IMG/pdf/AlsiratEMERGENZA\\_FINITA\\_DEFINITIVA.pdf](http://www.meltingpot.org/IMG/pdf/AlsiratEMERGENZA_FINITA_DEFINITIVA.pdf).
- NAPHT. 2014. National Action Plan against Human Trafficking (2014-2016). *Ministry for Women's Rights, Urban Policy, Youth, and Sports*. Available at <https://perma.cc/PE6C-ARGW>.
- Newey, W. K., & West, K. D. 1994. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**(4), 703–708.
- Okkerse, Liesbet. 2008. How to measure labour market effects of immigration: A review. *Journal of Economic Surveys*, **22**(1), 1–30.
- Palmisano, L., & Sagnet, Y. 2016. *Ghetto Italia. I braccianti stranieri tra capolarato e sfruttamento*. Fandango Editore.
- Penal Code. 2009. Law n. 94/2009 item 10bis on dispositions on public security: Entry and illegal stay in the State territory (15 July 2009). *Italian Penal Code, Gazzetta Ufficiale* 170.
- Peri, G. 2016. Immigrants, Productivity, and Labor Markets. *Journal of Economic Perspectives*, **30**(4), 3–30.
- Pesaran, M. H. 2004. General diagnostic tests for cross section dependence in panels. *Econometric Reviews*, **34**(6-10), 1089–1117.

- Polchi, Vladimiro. 2011. Immigrati, il 2011 anno record di arrivi: Aumentano le richieste d'asilo, piú 102% (30 December 2011). *La Repubblica*. Retrieved from <http://repubblica.it>.
- Rubin, D. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, **5**(5), 688–701.
- Sasso, Michele, & Sironi, Francesca. 2012. Scandalo profughi (12 October 2011). *L'Espresso*, Available at [http://www.meltingpot.org/IMG/pdf/Scandalo\\_profughi\\_ESPRESSO.pdf](http://www.meltingpot.org/IMG/pdf/Scandalo_profughi_ESPRESSO.pdf).
- Spagnolo, Chiara. 2017. Caporalato, in Salento 4 imprenditori condannati a 11 anni: Migranti schiavizzati (13 July 2017). *La Repubblica*. Retrieved from <http://bari.repubblica.it>.
- SPRAR. 2010. Rapporto annuale del sistema di protezione per richiedenti asilo e rifugiati. *Report 2010/2011 of the System for the Protection of Asylum Seekers and Refugees (SPRAR)*.
- SPRAR. 2011. Rapporto annuale del sistema di protezione per richiedenti asilo e rifugiati. *Report 2010/2011 of the System for the Protection of Asylum Seekers and Refugees (SPRAR)*.
- SPRAR. 2012. Rapporto annuale del sistema di protezione per richiedenti asilo e rifugiati. *Report 2011/2012 of the System for the Protection of Asylum Seekers and Refugees (SPRAR)*.
- Tumen, S. 2016. The Economic Impact of Syrian Refugees on Host Countries: Quasi-Experimental Evidence from Turkey. *American Economic Review*, **106**, 456–60.
- Vaiou, D., & Hadjimichalis, C. 1997. *With the sewing machine in the kitchen and the Poles in the fields. Cities, regions and informal work*. Athens: Exandas.
- Venturini, A. 1999. Do Immigrants Working Illegally Reduce the Natives' Legal Employment? Evidence from Italy. *Journal of Population Economics*, **12**(1), 135–154.
- Venturini, A., & Villosio, C. 2008. Labour-market assimilation of foreign workers in Italy. *Oxford Review of Economic Policy*, **24**(3), 517–541.
- Warren, R.E., & Passel, J. S. 1987. A count of the uncountable: Estimates of undocumented aliens counted in the 1980 United States Census. *Demography*, **24**(3), 375–393.
- WineSpectator. 2012. 2012 Vintage Reports. Retrieved from <https://winespectator.com>.
- Wooldridge, J. M. 2013. *Introductory Econometrics. A modern Approach*. 5 edn. Cengage Learning.

# The economic costs of hybrid wars: The case of Ukraine

with Julia Bluszczy\*

*Published in Defense and Peace Economics here.*

**Abstract** With more than ten thousand casualties, the ongoing hybrid Ukrainian war between pro-Russian separatists and the government in the Donbass region, Ukraine's productive core, has taken a severe toll on the country. Using cross-country panel data over the period 1995-2017, this paper estimates the causal effects of the Donbass war on Ukraine's GDP. Our counterfactual estimation by the synthetic control method shows that Ukraine's per capita GDP foregone due to the war amounts to 15.1% on average for 2013-2017. Separate analysis for the affected regions of Donetsk and Luhansk indicate an average causal effect of 47% for 2013-2016. Results are robust to pre-war confounds, namely, the Orange Revolution and Ukrainian-Russian gas disputes. As such, we discuss mechanisms underlying the war's causal effects on economic performance, which is of broader relevance for debates on the role of government in hybrid conflict management.

**Keywords** hybrid conflict, economic costs, Ukraine crisis, synthetic control, conflict management

**JEL Codes:** C21, F5, O41, O47

---

\*Humboldt University of Berlin, Germany. Email: [julia.bluszczy@gmail.com](mailto:julia.bluszczy@gmail.com).

## Introduction

The effectiveness of hybrid warfare, described as the use of insurgent tactics coupled with conventional military power to achieve politico-strategic goals, relies on ethnic grievances and weak civil societies (Lanoszka, 2016). Due to the close link between belligerents and the target society as well as the evolving and unpredictable nature of such conflicts, governments face several challenges in developing conflict management strategies (Giegerich, 2016). For instance, counter-insurgencies may deteriorate rather than improve the country's state of affairs.

Military conflicts always entail large costs, including economic, social, political, psychological and environmental ones. A vast literature is devoted to the ex-post evaluation of the economic costs of conflict to assess the losses incurred by the states and the civil society. Starting from Keynes (1919), many studies show that war has persistent negative consequences on the welfare of the populations involved (see, e.g. Gates, 2012; Koubi, 2005; Abadie and Gardeazabal, 2003). To our knowledge this is the first paper that identifies and quantifies the causal effects of the hybrid war in the Donbass region on Ukraine's GDP. Additionally, our study discusses mechanisms underlying these causal effects, and statistical challenges in the analysis arising from the complex nature of hybrid wars.

**The war.** The Donbass war is an armed conflict between anti-government groups of pro-Russian separatists and the Ukrainian government, taking place in the aftermath of the 2013 Euromaidan protests and the 2014 Ukrainian revolution. Thereby, this war embodies the hybrid form of state-on-state (Russian-Ukrainian) conflict. Located in eastern Ukraine, the Donbass region is considered Ukraine's productive core due to coal mining and highly productive heavy industry.<sup>1</sup> As such, the Donbass war has taken a severe toll on Ukraine, especially in terms of production, employment, number of displaced persons, and civilian as well as military casualties (Angelovski, 2015).

The Donbass is of considerable importance for Ukraine's production. Before the 2014 Ukrainian Revolution, this region accounted for about a quarter of the country's exports and more than 15% of capital investment (Ukrstat, 2014). For instance, the Donbass used to provide raw materials such as coal, steel and other industrial goods to international manufactur-

---

<sup>1</sup>For a map of the conflict, see Figures 9 and 10 in the Appendix 6.1.

ing industries. As of August 2014, the industrial production dropped by 60% and 85% in the Donbass regions of Donetsk and Luhansk, respectively, due to power cuts and the destruction of transport infrastructures (Havlik, 2014). Overall, major reasons for the decline of Ukraine’s economic activity are high costs of trade together with employment, agricultural and financial losses, compressed government spending, and the partial military mobilization coupled with growing political instability (Foreign Affairs Ministry, 2015).

As a *hybrid* and complex form of warfare, the Donbass war is an especially interesting case study.<sup>2</sup> Modern conflicts are indeed more likely to arise as a consequence of regional struggles with governments facing non-governmental actors who operate in concert with external players. Specifically, hybrid wars especially threaten the government’s sovereignty due to lack of soil governance and means to tackle issues like unclear front lines or friendly/enemy areas; unclear *casus belli* and politico-strategic goals; and new tactics that focus on the weakening of governments and state institutions rather than on direct combat (Deshpande, 2018).

In this respect, this paper aims to help deconstructing the complexity of the Ukrainian conflict by (i) providing formal statistical evidence on the causal effects on the country’s economy both at national and regional levels and (ii) discussing market mechanisms underlying these effects, also in the prospects of governments’ conflict management and resolution.

**The war’s outcomes.** Due to the Donbass’ strategic role in the country’s economy and its large contribution to the GDP, we expect the war to have a negative *causal* impact on this outcome. Although its components are of relative importance in determining the causal effect, we focus on the GDP foregone as an aggregate measure of the economic costs for two reasons. First, we want to allow for a higher degree of internal and external validity of our analysis. This approach is also followed by, e.g. Costalli et al. (2017), Horiuchi and Mayerson (2015), and Abadie and Gardeazabal (2003), who find strong significant average per capita GDP losses ranging

---

<sup>2</sup>As a matter of fact, despite more than ten thousand casualties and continuous fights (OHCHR, 2017), neither Ukraine nor any other entity declared the war status: the Ukrainian government referred to it as an anti-terrorist operation, and, on the other side, Russia admitted that intelligence military forces were sent to Ukraine, but denies the use of regular troops (Walker, 2015). As a result, although there are many signs indicating Russia’s involvement in the Donbass war (Rácz, 2015), the lack of undeniable confirmation from Kremlin’s side complicates the relationship between both countries and hinders any mitigation of the conflict.

from 8.6% to 17.5% (see, e.g. Gardeazabal, 2012, for a review of existing studies). As such, constructing an accurate (in terms of quality of the available data) and reliable (in terms of theoretical guarantees) counterfactual for Ukraine’s per capita GDP contributes to the cross-country comparison of our results with the literature’s. Second, since the Donbass war is still ongoing at the time of writing, it is difficult to give precise estimates of other types of costs due to lack of data. In light of these factors, we consider per capita GDP foregone as the main measure of welfare loss.

**Empirical strategy.** We use the Synthetic Control Method (SCM) to estimate causal effects of this war on Ukraine’s GDP per capita. Since its first application by Abadie and Gardeazabal (2003) and later formalisation in Abadie et al. (2010), this method has been more recently employed to estimate the causal effects of conflicts on GDP by, e.g. Echevarría and García-Enríquez (2019b,a) but also by, e.g. Albalade and Bel (2020) to estimate the effects of government formation deadlocks on GDP growth.

Building on the potential outcomes approach (Rubin, 1974), we obtain the counterfactual, ‘synthetic’, Ukraine as a weighted average of control (unaffected) countries with weights reflecting the resemblance of both the outcome variable and outcome predictors in Ukraine before the war’s outbreak. A country-level panel data over the period 1995-2017 is used for the analysis. Causal effects are estimated by computing the yearly difference in GDP per capita between Ukraine and its synthetic counterpart after the eruption of the war. Moreover, we apply the SCM iteratively to check for other potential shocks taking place in Ukraine before the Donbass war, in particular, the 2004 Orange Revolution, and the 2009 gas dispute with Russia. Finally, since the war is likely to affect the Ukrainian territory unequally, we further conduct a similar analysis for the Donbass regions of Donetsk and Luhansk.

**Preview of results.** Results indicate that due to the Donbass war, whose start is set to 2013, Ukraine’s foregone GDP per capita amounts to 15.1% on average in the post-war period and, respectively, 5.23% (\$460.26), 9.18% (\$832.96), 19.63% (\$1,823.78), 19.80% (\$1,893.38), 21.67% (\$2,184.13) in 2013, 2014, 2015, 2016, and 2017. The obtained estimates are validated by a series of robustness checks. After iteratively applying the SCM, we find that gas disputes led to an overestimation of the previous causal effects by 1.21 percentage points (\$128.04) on average. Instead, our findings show that the Orange Revolution did not considerably influence Ukraine’s



economic development and, thus, did not confound the obtained causal estimates of the war. Lastly, results from the regional analysis confirm the devastating effect of the war for the Donbass area. In particular, we estimate that Donetsk’s per capita Gross Regional Product (GRP) dropped by 42% (\$4,294) on average due to the war. Estimates for Luhansk are of even larger magnitude with a per capita GRP average decrease of 52% (\$3,355).

## 1 Empirical strategy

This section presents the SCM as developed by Abadie and Gardeazabal (2003) and later refined by Abadie et al. (2010). In addition to the identification and estimation strategy, we discuss advantages of the SCM as well as its limitations especially related to inference.

The true causal impact of a conflict on per capita GDP is given by outcome differences between Ukraine after the war and its counterfactual without the war. The SCM builds upon the potential outcomes approach (Rubin, 1974) to estimate this counterfactual, ‘synthetic’ Ukraine, by weighting units in the control group before the war to resemble Ukraine in all outcome-relevant variables, in particular observed time-varying covariates and a set of pre-intervention outcomes. Once the control group is weighted to predict Ukraine’s per capita GDP path before the war, post-war differences would only be due to the war if Ukraine’s per capita GDP is accurately fitted by the synthetic control pre-war. Our main parameter of interest is the Average Treatment effect on the Treated (ATT) over the periods after treatment, which can be computed as the post-war average difference between the observed outcome of Ukraine and synthetic Ukraine (Gobillon and Magnac, 2016; Abadie et al., 2010).

Consider  $i = 1, \dots, J + 1$  countries and  $t = 1, \dots, T$  time periods with  $1 \leq T_0 < T$  pre-war periods, and define  $Y_{it}^N$  to be the per capita GDP of Ukraine  $i = 1$  in time  $t$ , if not exposed to the war. Let Ukraine be the only recipient of the war, and let any other  $j = 2, \dots, J + 1$  country be unaffected by the conflict for a total of  $J$  unaffected units. Note that SCM assumes that no country anticipates the war’s outbreak before the time period  $T$ , and that there are no spillover effects of the conflict on the  $J$  control regions after the war (known as Stable Unit Treatment Value Assumption, SUTVA). We further denote the observed outcome for unit  $i$  at time  $t$  as  $Y_{it} = Y_{it}^N + \alpha_{it}D_{it}$  where  $D_{it}$  serves as a conflict indicator taking value 1 for Ukraine after

2012 and 0 otherwise. The war causal effect to be estimated is given by the Treatment effect on the Treated,  $TT_t = Y_{1t} - Y_{1t}^N$  for  $t > T_0$ , and the empirical challenge is to reconstruct the counterfactual  $Y_{1t}^N$ , i.e. the post-treatment outcome of the treated unit had it not been treated. Once the counterfactual outcome,  $\hat{Y}_{1t}^N$ , is estimated, the ATT over the  $T - T_0$  periods after treatment is computed as  $\hat{\alpha}_1 = \frac{1}{T - T_0} \sum_{t > T_0} (Y_{1t} - \hat{Y}_{1t}^N)$ .

Consider a  $(J + 1)$  vector of optimal weights  $W^* = (w_2^*, \dots, w_{J+1}^*)'$  with  $w_j \geq 0$  for  $j = 2, \dots, J + 1$  and  $w_2 + \dots + w_{J+1} = 1$  for  $J$  control units such that  $\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j^* Y_{jt}^N$ . The synthetic control method recreates this counterfactual with a convex combination of untreated units, i.e.  $\hat{Y}_{1,t > T_0}^N = \sum_{j=2}^{J+1} w_j^* Y_{j,t > T_0}^N$ . The aim of this analysis is to obtain the ATT over the periods after treatment defined as:

$$\hat{\alpha}_1 = \frac{1}{T - T_0} \sum_{t > T_0} [Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N]. \quad (1)$$

The estimation of the optimal  $W^*$  follows a nested optimization procedure. First, an inner optimization minimizes the Euclidean distance between  $X_1$  and  $X_0 W$ ,  $(r + k) \times 1$  and  $(r + k) \times (J)$  matrices, respectively, containing  $k$  covariates and  $r$  linear combinations of pre-war outcomes used as predictors (2):

$$W^* = \arg \min_W \|X_1 - X_0 W\| v = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}, \quad (2)$$

where  $V$  is a  $(r + k) \times (r + k)$  symmetric diagonal matrix with non-negative components, in which the diagonal elements  $v = (v_1, \dots, v_{r+k})$  are the predictor weights assigned to the fitted pre-intervention variables. In an outer optimization,  $V^*$  can be estimated such that the Mean Squared Error (MSE) of per-capita GDP outcomes is minimized for pre-treatment periods according to  $V^* = \arg \min_V (Y_1 - Y_0 W^*(V))' (Y_1 - Y_0 W^*(V))$ , where  $Y_1$  denotes pre-war outcomes of Ukraine and  $Y_0$  refers to linear combinations of pre-war outcomes of control countries, which can be, e.g. averaged over some pre-war periods.

The SCM identifying assumptions are twofold. First, the outcome of all countries is required to follow a linear model, like, e.g. a factor model including interactive fixed effects that capture time-varying unobserved heterogeneity (see Abadie et al., 2010, and Ahn et al., 2013). Since Abadie and

Gardeazabal (2003) introduce the SCM using GDP as the dependent variable, we consider this assumption as fulfilled (see, e.g. Costalli et al., 2017; Horiuchi and Mayerson, 2015 for similar choices). Second, there exist optimal (non-negative) weights (smaller than and adding up to one) that build the synthetic control as a convex linear combination of control countries matching a set of covariates and outcomes pre-war. This is violated in the presence of interpolation bias, i.e. if the synthetic control obtains weights for countries that largely differ in terms of unobservable confounders that may trigger any change in the outcome. In our context, confounders could consist of, e.g. unobserved time-varying factors driving both GDP and the conflict. To avoid the interpolation bias we restrict the control group to countries of the former Soviet Union and Eastern Bloc which are not treated with the Donbass war or other shocks, as they most accurately reflect the Ukrainian economy, and we exclude Russia because it is part of the Donbass war, of the 2014 annexation of Crimea, and the consequent economic sanctions imposed by the European Union and the United States. Therefore, provided that the number of pre-war periods is large and interpolation bias is not present, the synthetic control approximately fits Ukraine also in its individual time-varying heterogeneity (Abadie et al., 2010). In such cases, the SCM provides unbiased estimates of the counterfactual with more identification power than traditional regression methods accounting only for time-invariant unobserved differences (Gobillon and Magnac, 2016).

We perform standard statistical inference and robustness analyses by (a) in-space placebo tests, and also (b) in-time tests. In the latter case (b), we apply the SCM on Ukraine’s outcome pre-war as a means to uncover likely confounding effects of two events: the 2009 gas disputes with Russia, and the 2004 Orange Revolution. In the former case (a), we build a synthetic control for each country in the control group, and we estimate the corresponding ATT (Abadie et al., 2010). We compute empirical p-values as the probability to obtain ATTs at least as large as the treated unit’s (in absolute terms), i.e.  $\frac{\sum_{i=1}^{J+1} \mathbf{1}(|ATT_i| \geq |ATT_1|)}{J+1}$  for  $i = 1, \dots, J+1$  where  $\mathbf{1}(|ATT_i| \geq |ATT_1|)$  takes value 1 when  $|ATT_i|$  is equal or larger than  $|ATT_1|$  and 0 otherwise.<sup>3</sup>

Additionally, we perform in-space placebo tests accounting for the prediction accuracy of the estimated synthetics in the pre-war period. For every country  $i = 1, \dots, J+1$ , we first compute the pre-war average Mean Predic-

---

<sup>3</sup>When the magnitude of  $ATT_1$  is extreme relative to the permutation distribution, we obtain the smallest p-value of size  $\frac{1}{J+1}$ .

tion Error (MPE) defined as  $\frac{1}{T_0} \sum_{t \leq T_0} [Y_{it} - \hat{Y}_{it}^N]$ , with  $\hat{Y}_{it}^N$  being the synthetic control estimated for every country in the sample. Second, we condition on the MPE to obtain empirical p-values as  $\frac{\sum_{i=1}^{J+1} \mathbf{1}(|ATT_i| \geq |ATT_1|)}{J+1}$  s.t.  $|MPE_i| \leq |MPE_1|$ . Note that this test differs from the inferential technique performed by, e.g. Abadie et al. (2010) who compute the ratio between post-intervention and pre-intervention MSE. In this case, the numerator results to be inflated in the presence of, for instance, a large causal effect in one single post-war period, as squaring post-war gaps assigns a higher weight to exceptionally large deviations. On the contrary, a counterweight of this effect in the denominator for pre-intervention MSE is unlikely to occur as every placebo country with a much (typically, five to three times) higher MSE than the one of the treated unit is excluded from the computation of the p-values. Motivated by the above as well as by recommendations in Arkhangelsky et al. (2019), Ferman and Pinto (2017), and Firpo and Possebom (2018), MPE-based inference is also presented.

What are the advantages of the SCM over other techniques in our context? Due to the variety of costs that can be attributed to armed conflicts, researchers have adopted three types of evaluation tools: cost-accounting, regression-based, and counterfactual methods (for a detailed overview see, e.g. Gardeazabal, 2012; de Groot et al., 2009). In this paper, we find the SCM to be most suited to evaluate the cost of the Donbass war for at least five reasons. First, unlike cost-accounting methods employed in, e.g. Skaperdas et al. (2009), Bilmes and Stiglitz (2006) and Arunatilake et al. (2001), the SCM does not require multiple calculations of a broad variety of costs, which relies on the availability and quality of governmental data as well as on expertise in listing all types of costs and avoiding double counting. In addition, the SCM allows to perform statistical inference and assess the uncertainty inherent in the cost estimate. Second, in contrast to panel data and time-series methods used by, e.g. Enders et al. (1992) and Barro and Lee (1994), the SCM is more robust against the presence of unaccounted-for factors that may influence the outcome. By assuming a factor model specification, the SCM allows for a multidimensional unobserved heterogeneity, i.e. for multiple interactive effects, not just additive ones as imposed, e.g. in the difference-in-differences setting (Gobillon and Magnac, 2016). In practice, interactive effects can be considered as time-varying fixed effects like, for example, country-specific variations in strategic alliances. Therefore, the SCM generalizes the difference-in-differences method allow-

ing to clearly identify the causal effect of the Donbass war on GDP per capita, disentangling the causal effect from other unobserved time-varying confounders. Third, the SCM improves upon other regression methods because it performs well with small-sized groups, it safeguards against ad-hoc model specification searches, and precludes negative weights, thus, avoids extrapolation outside the support of the data (Abadie, 2019). Finally, the SCM estimates sparse weights for the control units allowing to assess their contribution to the counterfactual, and to evaluate directions of potential biases. As a result, the SCM is a well established causal inference tool, according to Athey and Imbens (2017) one of the most important innovations in the evaluation literature in the last fifteen years.

## 2 Data

We use yearly country-level panel data over the period 1995-2017 obtained from the World Development Indicators database of the World Bank. The dependent variable used in the SCM analysis is the GDP per capita (GDPpc) in 2011 dollars (PPP). Further, outcome predictors used to match Ukraine in the pre-war period are chosen based on literature review (e.g. Abadie and Gardeazabal, 2003). We include inflation measured by consumer price index due to the prevalence of hyperinflation in post-Soviet states, and its influence on economic development. Further, we control for domestic investment with gross fixed capital formation (GFCF) as a percentage of GDP, and we measure the dependence on trade with Russia as the sum of share of exports and imports with the Russian Federation in countries' total international trade (TradeDep). Finally, to account for political and socio-economic resemblance, we also include the Human Development Index (HDI) which is a composite indicator of life expectancy, education, and per capita income, as well as the Polity variable from the Polity IV project dataset in which values equal to 10 (-10) indicate a strongly democratic (autocratic) regime (Marshall, 2017). In the SCM estimation, we match on covariates' averages over the 1995-2012 period, and on two outcome lags. The following Table 1 provides data descriptive statistics, while the variables' full description can be found in Table C.1 in the Appendix 6.3.

Table 1. Descriptive statistics of variables.

|           | Mean     | Standard Deviation | Maximum  | Minimum |
|-----------|----------|--------------------|----------|---------|
| TradeDep  | 0.15     | 0.15               | 0.76     | 0.01    |
| GFCF      | 23.79    | 6.37               | 57.71    | 5.39    |
| GDPpc     | 14032.00 | 8509.13            | 31339.00 | 1043.00 |
| Inflation | 18.45    | 59.98              | 1058.00  | -8.52   |
| Polity    | 4.96     | 6.25               | 10       | -8      |
| HDI       | 0.74     | 0.08               | 0.81     | 0.53    |

The SCM makes a crucial assumption that Ukraine’s GDP per capita and all its predictors have to lie within the convex hull spanned by the countries from the donor pool, such that a convex combination of the control countries can actually resemble the treated unit. Figure 1 shows evidence on the presence of such common support. The plot presents boxplots of all predictor variables after their mean normalization, i.e. for each variable we computed  $(X_k - \mu_k)$  for  $k = 1, \dots, 6$ . It can be inferred from the the plot that the values of the predictors for Ukraine lie within the spectrum spanned by the units from the donor pool.

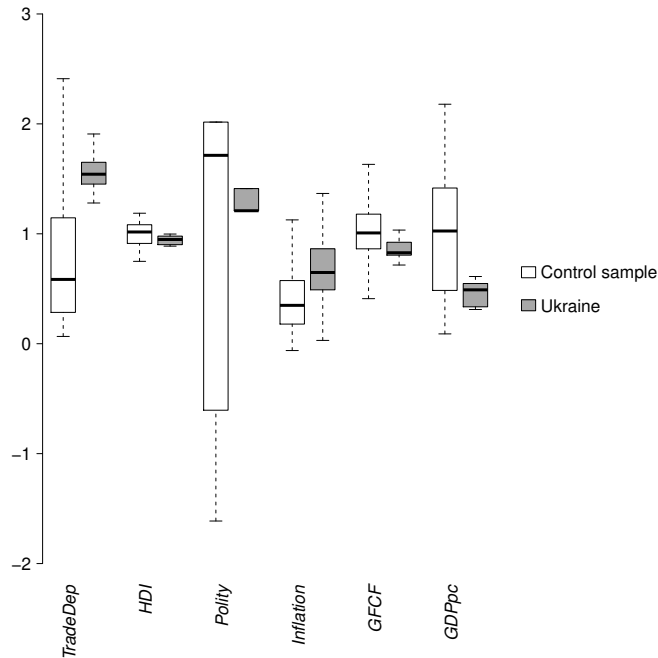


Figure 1. Boxplots of mean-corrected variables.

As outlined in the Data, synthetic Ukraine is built as a weighted average of former Soviet Union and Eastern Bloc countries to most accurately resemble its unobserved fiscal and economic conditions over time. We also excluded countries experiencing other shocks in the considered pre-war period. As a result, the control group comprises 17 countries, which are listed

in Table 2 of the Results.

A last note regarding the onset of the Donbass war. The war burst out in 2014, however, it was preceded by the 2013 violent Euromaidan protests and a period of high political instability. For this reason, we assign the year 2013 as the start of the war. Consequently, we estimate the counterfactual over 18 pre-war periods, and we predict the outcome over five post-war periods. As specified in the Data, a precise and robust fit between actual and counterfactual outcome over the whole pre-war period is necessary to guarantee the validity of the counterfactual estimate itself.

### 3 Results

Using the SCM, we first show how synthetic Ukraine fits Ukraine’s GDP per capita before the war to provide an unbiased counterfactual after the war, and we compute causal effects. Second, we assess statistical significance by placebo tests, and we perform a set of confoundedness as well as sparsity checks. Third, using analogous analyses, we provide further evidence on the war’s causal effects for Ukraine’s most affected regions.<sup>4</sup>

Table 2 shows that synthetic Ukraine is best reconstructed as a weighted average of four countries, namely, Armenia, Bulgaria, Moldova, and Slovenia - with Moldova and Armenia yielding the highest weights.

Table 2. Ukraine’s control sample with corresponding weights.

| Country         | Weight | Country         | Weight |
|-----------------|--------|-----------------|--------|
| Armenia         | 0.333  | Latvia          | 0      |
| Azerbaijan      | 0      | Lithuania       | 0      |
| Belarus         | 0      | Moldova         | 0.452  |
| Bulgaria        | 0.152  | Poland          | 0      |
| Czech Republic  | 0      | Romania         | 0      |
| Estonia         | 0      | Slovak Republic | 0      |
| Hungary         | 0      | Slovenia        | 0.063  |
| Kazakhstan      | 0      | Tajikistan      | 0      |
| Kyrgyz Republic | 0      |                 |        |

Furthermore, Table 3 displays the results of the estimation, and shows that synthetic Ukraine accurately reproduces mean values of the covariates before the war. As a measure of overall goodness of fit, Table 3 reports the Mean of the Absolute Prediction Errors (MAPE) which amounts to

<sup>4</sup>We use the statistical software R and, particularly, the Synth package (Abadie et al., 2011).

4% relative to the mean value of Ukraine's per capita GDP in the pre-war period.

Table 3. Outcome predictor means and weights.

| Covariate   | Weight | Ukraine |           | Donor pool |
|-------------|--------|---------|-----------|------------|
|             |        | Real    | Synthetic |            |
| Inflation   | 0.004  | 36.44   | 23.04     | 20.72      |
| GFCF        | 0.032  | 20.91   | 22.89     | 24.17      |
| TradeDep    | 0.001  | 0.23    | 0.19      | 0.14       |
| HDI         | 0.001  | 0.70    | 0.68      | 0.74       |
| Polity      | 0.001  | 6.50    | 6.83      | 4.78       |
| GDPpc(2000) | 0.577  | 4797.38 | 4797.03   | 10650.16   |
| GDPpc(2012) | 0.385  | 8322.17 | 8538.05   | 17963.32   |
| MAPE        |        |         | 0.04      |            |

*Note:* All variables are averaged for the 1995-2012 period except for lagged values of GDP per capita.

Figure 2 displays the trends of per capita GDP of Ukraine and its synthetic counterpart. It clearly shows that both follow a very similar path until 2012 and deviate considerably afterwards. The ATT - computed as the post-war average difference between observed and synthetic GDP per capita - amounts to 15.1%. In particular, yearly differences equal to, respectively, 5.23% (\$460.26), 9.18% (\$832.96), 19.63% (\$1,823.78), 19.80% (\$1,893.38), and 21.67% (\$2,184.13) in 2013, 2014, 2015, 2016, and 2017.



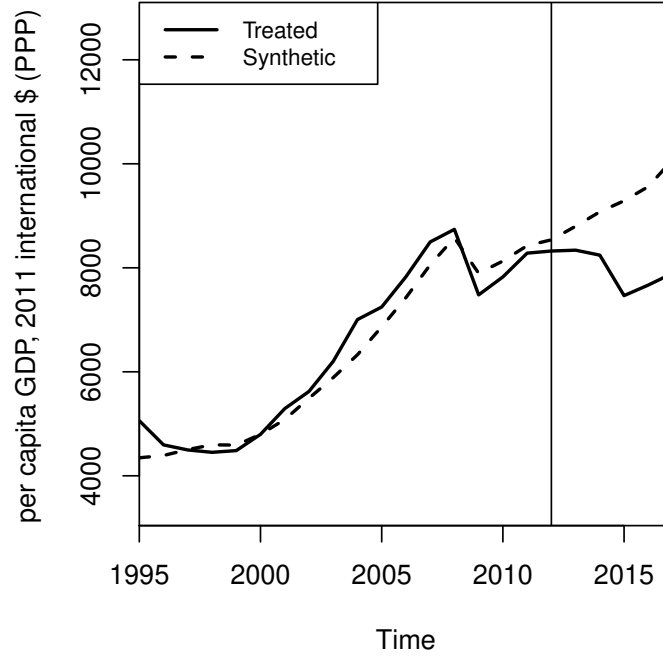


Figure 2. Trends in GDP per capita: Ukraine vs. synthetic Ukraine.

The SCM does not allow for usual large sample inferential techniques. Instead, it provides a framework for placebo tests. Figure 3 shows the graphical representation of placebo tests for Ukraine and control countries. To be conservative, we exclude from the test control countries with MSE five times higher than the one obtained for Ukraine (as suggested by Abadie et al., 2010). As a result, we exclude six countries, i.e. Azerbaijan, Czech Republic, Estonia, Poland, Slovenia, and Tajikistan from the plot, which leaves 11 remaining control countries.

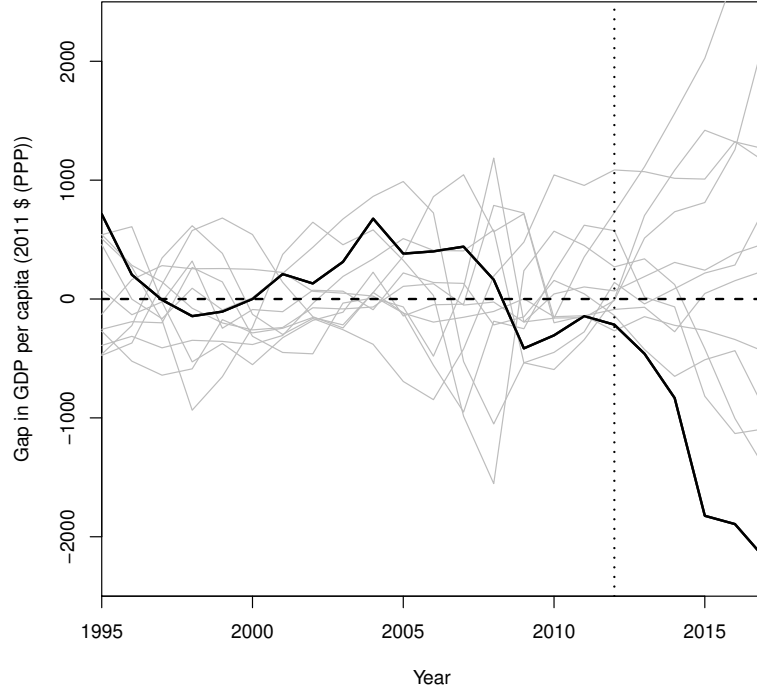


Figure 3. Gaps in GDP per capita in Ukraine and placebo gaps. Countries with poor pre-war fit excluded.

Based on placebo tests reported in Figure 3, we conclude that the ATT estimated for Ukraine is larger in magnitude than all placebo ATTs, leading to an empirical p-value of 8% (one over 12) which is the lowest level that we can reach given the size of the considered control group.

Additionally, we further assess the statistical significance of the ATT by computing post-/pre-war MSE ratios. Differently from the test above, this statistic accounts for the goodness of fit of the placebos before the war, and is therefore computed for all 18 units. Figure 4 shows that Ukraine presents the second biggest ratio, yielding a statistical significance level of 11% (two over 18). Finally, as discussed in the Empirical strategy, since the MSE criterion overweights large gaps, the ATT of each unit is also plotted, for robustness, against its pre-war Mean Prediction Error (MPE), the mean gap before treatment.<sup>5</sup> Figure 5 plots the ATT-MPE test. Since no control unit lies in the highlighted area showing a more extreme ATT as well as a smaller MPE than Ukraine (in absolute terms), the ATT-MPE test suggests

<sup>5</sup>Since this test explicitly accounts for the prediction accuracy of all synthetics in the pre-war period, we do not exclude control countries with MSE five times higher than the one obtained for Ukraine.

a significant average causal effect at the 5% level (one over 18).

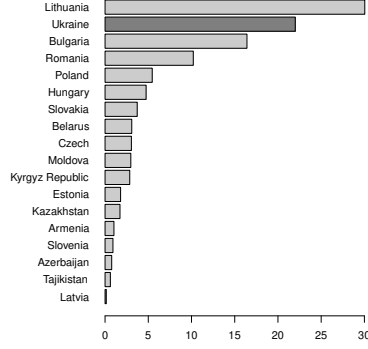


Figure 4. Ratio of mean squared prediction error in post- and pre-war periods.

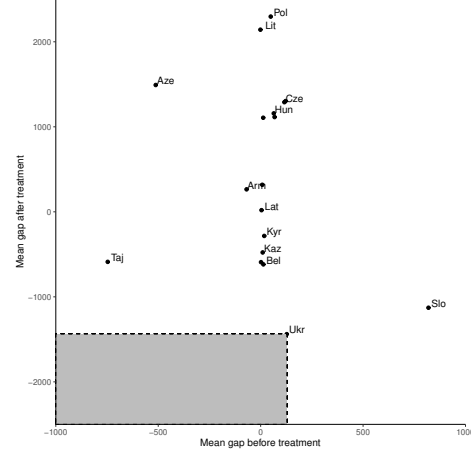


Figure 5. Mean prediction error in post- and pre-war periods (ATT vs. MPE).

### 3.1 Robustness and sensitivity

To check whether the estimated synthetic Ukraine is robust against different linear combinations of country weights, we perform a Leave-One-Out (LOO) estimation. This means we iteratively build synthetic Ukraine excluding one control unit at the time among those units with positive weights in the synthetic Ukraine (i.e. Slovenia, Bulgaria, Armenia, and Moldova). Figure 11 in the Appendix 6.2 shows that the LOO synthetic controls closely match the original synthetic Ukraine that includes all control countries, verifying the robustness of the original synthetic. As a result, no country is found to be pivotal to the results. Further, since each LOO synthetic control is above our original estimate, our results can be interpreted as a lower-bound for the war's causal effects on Ukraine's GDP per capita. In particular, excluding Moldova (Armenia), the country receiving the (second) highest weight in the original estimation, would cause higher estimates by 10% (20%) on average in the post-war period. Table 4 reports all LOO causal effect estimates on Ukraine's per capita GDP in each post-war period and on average over the post-war period (ATT).

Table 4. LOO estimation: Ukraine’s per capita GDP losses causal to the Donbass war in each post-war year, and on average in the post-war period (ATT).

| Year | Slovenia | Bulgaria | Armenia | Moldova | Original |
|------|----------|----------|---------|---------|----------|
| 2013 | 490.8    | 649.0    | 686.6   | 588.7   | 460.3    |
| 2014 | 914.9    | 1083.7   | 1038.7  | 866.9   | 833.0    |
| 2015 | 1932.3   | 2040.6   | 2089.1  | 1982.8  | 1823.8   |
| 2016 | 1931.0   | 2103.1   | 2321.3  | 2146.8  | 1893.4   |
| 2017 | 2236.1   | 2455.7   | 2553.9  | 2351.6  | 2184.1   |
| ATT  | 1501.0   | 1666.4   | 1737.9  | 1587.4  | 1438.9   |

*Note:* GDP in 2011 international dollars, PPP.

The presence of substantial spillovers would violate the no interference assumption (SUTVA) and lead to a bias in the synthetic control estimate. This may be especially the case for Ukraine’s neighboring countries such as Moldova (see map in Figure 10 in the Appendix 6.1). However, awareness of the type of potential spillovers allows to evaluate the validity of the counterfactual estimate and the directions of potential biases (Abadie, 2019). For instance, if the economic growth of neighboring countries was negatively affected by the Donbass conflict in the post-war period (perhaps because they diverted demand and investment from Ukraine) this would imply that the synthetic control represents a lower bound on the magnitude of the war’s negative effects on Ukraine’s GDP. We follow Cao and Dowd (2019) to gauge the potential effects of spillovers on our estimates. The test uses all 18 pre-war GDP per capita values as predictors, and considers Ukraine’s neighboring countries of Belarus, Moldova, Hungary, Poland, Romania, and Slovak Republic as the potentially affected units. Results confirm that our estimates are a lower bound of the war’s effect, and indicates that accounting for spillovers would lead to higher GDP losses of 1.8% in 2015, 15% in 2016, and 20% in 2017.<sup>6</sup> The Discussion Section outlines possible channels through which the Donbass war may (or not) have affected the per capita GDP of countries in the control group.

Moreover, we perform confoundedness checks obtained by iteratively estimating synthetic controls on the pre-war period to account for previous Ukrainian-Russian disputes. This allows to verify if Ukraine’s exposure to other external shocks also affects its outcome path. In particular, we analyze two events: the 2009 gas disputes with Russia, and the 2004 Orange Revolution. For this purpose, we iterate the SCM moving the treatment

<sup>6</sup>We thank an anonymous referee for suggesting this test.

period  $T_0 + 1$  to 2009 and 2004, respectively. In this case, we calculate treatment effects only until  $T = 2012$  in order to exclude effects of the Donbass war and obtain 'pure' effects of these events on Ukrainian economy.

The Orange Revolution is a series of political protests leading to a period of political instability that could have caused a slowdown in Ukraine's GDP. However, estimation results show that the ATT is not statistically significant at conventional levels, namely, ten over 18 units have a larger ATT (in magnitude) than Ukraine (p-value=0.55). This is displayed in Figures 12 and 13 in the Appendix 6.2.

Russian-Ukrainian gas disputes in 2009 represented a trade conflict over gas prices and their terms of export. Since no agreement was reached, Russia interrupted gas supplies to Ukraine which served as a transit country for Europe. Counterfactual estimation reveals a gap between observed and synthetic outcomes in 2009, indicating that Ukraine's GDP per capita would have been higher without the shock. This is shown in Figure 15 in the Appendix 6.2. In particular, gas disputes cause a one-time outcome level change in 2009 after which trends are parallel again. Placebo tests indicate that causal effects are significant at 10% level. Yet, these estimates may be confounded by two factors. First, gas prices increased after 2009, potentially affecting the GDP of control countries. This would lead to a bias in the SC due to violation of the no spillover assumption (SUTVA). Second, the effects of the 2009 financial crisis cannot be disentangled from those of the gas disputes.

Although the 2009 gap may not be entirely attributable to the gas disputes, we compute the confounding effects of the shock on the causal effects of the Donbass war. Being fairly constant for 2009-2012, outcome gaps caused by gas disputes ( $\hat{\Delta}_{gas}$ ) are subtracted from the per capita GDP values of synthetic Ukraine in consecutive years, and the Donbass war's causal effects are adjusted as shown in Table 5<sup>7</sup>. Clearly, accounting for gas disputes caused the loss to decrease from original yearly average of \$1,438.9 to \$1,310.86. These results, also displayed in Figure 14 in the Appendix 6.2, suggest that the 2009 events led to the overestimation of causal effects of the conflict by an average of 1.21 percentage points (\$128.04). As a result, the lower-bound for Ukraine's per capita GDP foregone due to the war amounts

---

<sup>7</sup>Causal effects of the gas disputes are computed as  $\hat{\Delta}_{gas} = 1/4 \sum_{t=2009}^{2012} (Y_{1t} - \hat{Y}_{gas,t}^N)$ , and the corrected counterfactual outcome accounting for such shock is  $\hat{Y}_{1,t}^{N,new} = \hat{Y}_{1t}^N - \hat{\Delta}_{gas}$  for  $t > 2012$ .

to 13.89%.<sup>8</sup>

Table 5. Per capita GDP differences between Ukraine and its synthetic control including effects of 2009 gas disputes with Russia.

| Year<br>t | GDP<br>$\hat{Y}_{1t}^N$ | Loss<br>$\hat{Y}_{1t}^N - Y_{1t}$ | GDP<br>$\hat{Y}_{1,t}^{N,new}$ | New Loss<br>$\hat{Y}_{1,t}^{N,new} - Y_{1t}$ | Diff.<br>in losses |
|-----------|-------------------------|-----------------------------------|--------------------------------|--|--------------------|
| 2013      | 8799.18                 | 460.26                            | 8647.00                        | 308.08                                       | 152.17             |
| 2014      | 9076.43                 | 832.96                            | 8912.73                        | 669.26                                       | 163.70             |
| 2015      | 9280.71                 | 1823.78                           | 9146.79                        | 1681.85                                      | 141.93             |
| 2016      | 9561.48                 | 1893.38                           | 9473.75                        | 1805.65                                      | 87.73              |
| 2017      | 10078.52                | 2184.13                           | 9983.85                        | 2089.46                                      | 94.67              |
| Average   | 9359.26                 | 1438.90 (15.10%)                  | 9231.21                        | 1310.86 (13.89%)                             | 128.04             |

*Note:* GDP in 2011 international dollars, PPP.

### 3.2 Regional synthetic control estimates

Since the Donbass war is limited to the territory of only two out of 24 Ukrainian regions (see map in Figure 9 in the Appendix 6.1), we estimate the impact of the war on the respective Gross Regional Product (GRP). Results from this estimation would also serve as a reliability check for the causal effects obtained at country level.

Regional SCM estimates are obtained with data from the State Statistics Service of Ukraine, and the first available period is 2004. Limited data availability constrains the choice of the variables included. These are exports of commodities (as a share of GRP), capital investments (as a share of GRP), unemployment rate, and per capita GRP. A detailed description of the data is to find in Table C.2 in the Appendix 6.3. The values of regions' GRP and capital investment are transformed from the Ukrainian currency (UAH) into international 2011 dollars using the exchange rates given in Table C.7 in the Appendix 6.3. The control group includes 22 Ukrainian regions with the exclusion of Kyiv City because its economy differs considerably from those of the other regions.<sup>9</sup> The complete list of the control units can be found in Table C.3 in the Appendix 6.3.

For the estimation, we use regional panel data from 2004 to 2016. As for the country-level estimation, the treatment is assigned in 2013 to account for

<sup>8</sup>These results are robust to an alternative correction strategy, namely, estimate Ukraine's SC and ATT after correcting Ukraine's outcomes for the causal effects of the disputes.

<sup>9</sup>Kyiv City is Ukraine's capital and its biggest, most affluent agglomeration. It accounts for nearly a quarter of the country's capital investment and its GRP per capita is roughly three times higher than Ukraine's average.

anticipation effects. We suspect that although armed conflicts did not start before 2014, there might have been regional tensions and hostilities that influenced social and economic living conditions of the local population. As a result, we match on nine pre-war periods to predict four post-war periods.

Tables C.3 and C.6 in the Appendix report the estimated unit and predictor weights, respectively, while Tables C.4 and C.5 show average values of covariates for the Donbass regions, their synthetic counterparts, and the whole control sample mean. It can be inferred that weighted averages accurately reconstruct all the outcome-relevant characteristics of the affected units.

Finally, Figures 6 and 7 plot the values of GRP per capita for Donetsk and Luhansk along with their synthetic counterparts. While in both cases observed and synthetic outcomes follow almost an identical trend until 2012, observed outcomes severely drop post-war. This estimation shows that, due to the Donbass war, Donetsk's and Luhansk's average GRP for 2013-2016 decreased by 42% (\$4,294) and 52% (\$3,355), respectively.

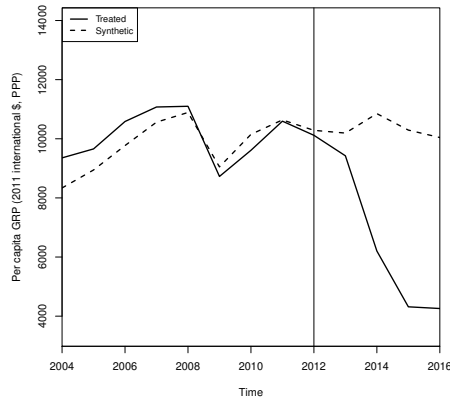


Figure 6. Per capita GRP trends in Donetsk vs. its synthetic counterpart.

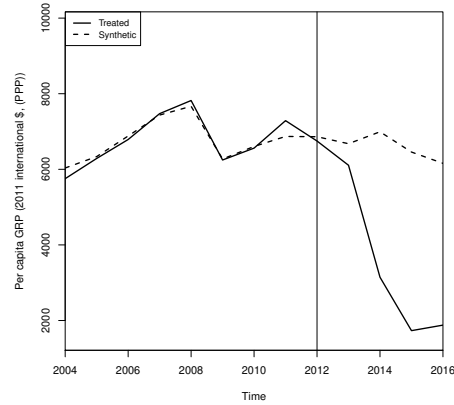


Figure 7. Per capita GRP trends in Luhansk vs. its synthetic counterpart.

Statistical significance of these causal estimates is confirmed by a series of placebo tests as shown in Figure 8 below, and Figures 16 and 17 in the Appendix 6.3.

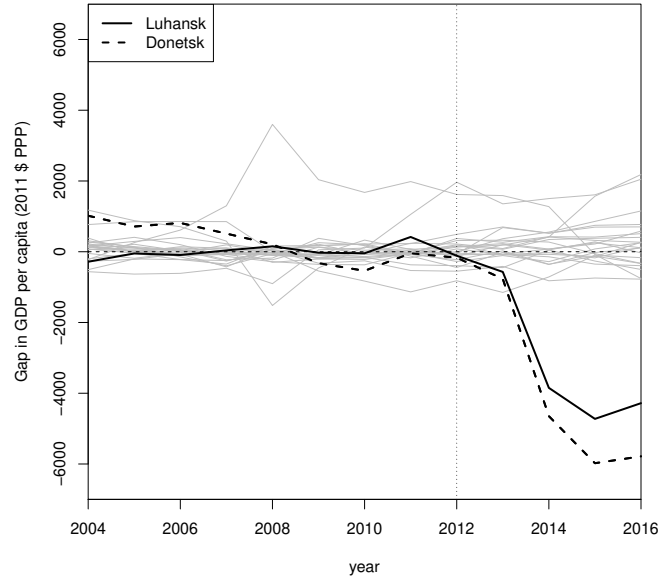


Figure 8. Gaps in per capita GRP in Donetsk, Luhansk and placebo gaps.

Compared to the country-level estimation, assumptions for the regional case are weaker, in particular, SUTVA. However, obtaining plausible causal estimates seems likely because Donetsk and Luhansk are the only regions directly affected by the fights. Following Cao and Dowd (2019), we test for spillovers in the potentially affected neighboring regions of Dnipropetrovsk, Kharkiv, and Zaporizhzhya for Donetsk, and Kharkiv for Luhansk (see map in Figure 9 in the Appendix 6.1) using all 9 pre-war GDP per capita values as predictors. We find that adjusting our estimates to account for spillovers would increase the magnitude of the ATT by 13% for Donetsk and 11% for Luhansk. Thus, similarly to the country-level estimation, the test shows that our estimates are a lower bound of the war’s causal effects.

## 4 Discussion

Our results from the counterfactual estimation show that Ukraine’s per capita GDP experienced a significant decline due to the Donbass war and, in its absence, would have instead followed a rather stable, slowly increasing trend. Importantly, we estimate that the gap between what actually happened in the Ukraine’s economy and its counterfactual widens over the four years after the official start of the war in 2014.



Although our study does not aim to statistically identify the particular GDP components that contributed to its decline, we hereby discuss possible mechanisms underlying the causal impacts of the Donbass war, and provide external anecdotal and formal evidence that supports our results. Additionally, we discuss the role of (I) economic sanctions and (II) military expenditures together with our study's limitations.

#### **4.1 Mechanisms underlying the GDP decline causal to the war**

There are several explanatory factors for the negative causal effects of the Donbass war on Ukraine's GDP per capita. These are, for instance, disruption to production, trade and employment, agricultural and financial losses, compression of public expenditures, and a partial military mobilization coupled with growing political instability.

Since its outbreak, the war took a heavy toll on Ukraine's economy especially due to trade disruptions and diverted government spending. Between May and September 2014, the European Bank for Reconstruction and Development adjusted its predictions on Ukraine's GDP loss from seven to nine percent (Reuters, 2014). In fact, only few months after the war's beginning, most of eastern Ukraine's core industries and infrastructure has been either shutdown or destroyed. Also, physical losses as well as water and electricity cuts made large facilities such as coal and power stations inoperable. For instance, coal mines have flooded because power cuts prevented water extraction, and the fewer transfers of the Donbass' exports of coal, steel and other industrial goods became more costly due to the destruction of bridges, roads and railways. As a result, during the war, the Donbass - previously Ukraine's largest industrial center - drastically reduced its contributions to the country's economy.

Several institutions and governmental sources reported estimates of the ongoing economic costs of the Donbass war. On September 2014, the Prime Minister Arseniy Yatsenyuk reported that the Donbass war was costing the government 6.15 million dollars a day (Wilson, 2015). Donetsk's politician and economist Alex Ryabchyn referred to a variety of reasons for such costs (Kuznetsov, 2014). He reported that (i) several manufacturing industries depend on raw materials coming from the Donbass, (ii) exports of coal, steel, and power heavily depend on the Donbass' output and transportation infrastructure, and (iii) one third of Ukraine's market is fostered by trade

with Russia. In the latter respect, the index of merchandise exports between Ukraine and Russia has fallen by 80% over 2013-2016 (Hornish, 2019).

Moreover, investors usually shy away in the presence of conflicts, and the confidence in the overall economy plummets. In the Donbass case, real foreign direct investment stagnated at about one percent of GDP on average over 2014-2018 (Åslund, 2019).

Furthermore, Ukraine's energy sector was also largely damaged by the failure of the Minsk peace process and the trade blockade against the Donbass started in 2015 and culminated in 2017. The latter was operated first by pro-Ukrainian activists, and later supported by the Ukrainian National Security Council, to block the transport of raw materials including, in particular, anthracite coal from the Donbass to the rest of Ukraine. This resulted in major production breaks for companies not only in eastern but also in western Donbass due to the interconnection of their production cycles (for more details see, e.g. Kostanyan and Remizov, 2017).

As an additional consequence of the blockade, Ukraine scored higher in terms of internal conflicts fought, and growing political instability which proxies for the propensity of a government collapse (IEP, 2018). Especially if coupled with increased internal conflicts, political instability has been shown to damage a country's economy by reducing investment, lowering the rates of productivity growth and, to a smaller extent, of physical and human capital accumulation (Hussain, 2014).

Yet, important questions remain open, namely, whether the negative effects of the Donbass war on Ukraine's economy have been reinforced by the government's mismanagement of the conflict and a weak external environment during the war. The latter includes, for instance, lower global commodity prices that resulted in a deterioration of Ukraine's terms of trade.

On the governmental side, instead, the World Bank reports that structural imbalances such as an already consolidated fiscal deficit were negatively adjusted in response to the war shock (World Bank, 2017). This likely resulted in a compression of domestic demand, increased public and guaranteed debt, and severe currency depreciation which induced deposit outflows, rising levels of nonperforming loans, and large numbers of bank failures, further reducing confidence in the economy. In this respect, an empirical study by Kochnev (2019b) estimates a nonlinear and on average positive association between Ukraine's stock market performance during the

war and investors' expectations on the prospect of conflict resolution.

In light of all the above, we can expect that not only the Donbass but also other Ukrainian regions were damaged by the conflict, in which case our estimated average GDP decline (43% for Donetsk and 52% for Luhansk) would represent an underestimate of the true causal effects for the Donbass region. Interestingly, using a difference-in-differences design and luminosity data to proxy for economic development, Kochnev (2019a) reports similar estimates, namely, a decrease in economic activity by 38% and 51% in the Donetsk and Luhansk regions, respectively.

## 4.2 Economic sanctions and military expenditures

One main limitation to our study is the possible violation of the SCM identifying Stable Unit Treatment Value Assumption (SUTVA). In particular, SUTVA requires the absence of spillover effects of the Donbass war, the treatment, on the GDP of countries and regions in the control group. If SUTVA is violated, the synthetic control built as a linear combination of untreated, though spilled over outcomes would be a biased estimate of the counterfactual in the post-war period. We identify two main unaccounted-for factors that may cause such spillovers, namely economic sanctions and military expenditures.

Concerning the first, we refer to the economic sanctions imposed on the Russian Federation by the EU in the aftermath of Crimea's annexation in 2014 and the Donbass war. These sanctions may have had an impact not only on the Russian Federation but also on the EU, especially because target and sender countries are economically interdependent and have cooperative political relations (Kaempfer and Lowenber, 2007; Drezner, 1999).<sup>10</sup> In light of this, our concern is that the GDP of countries and regions in the control group could be impacted, although indirectly, by these sanctions. However, we argue that this is likely a minor issue for at least three reasons. First, while sanctions affect exports directly, they only indirectly impact the GDP which is more sensitive to other dynamics such as taxation and countries' overall performance (Giumelli, 2017). Second, sanctions' effects are shown to be strongest in the very short-run, which indicates that GDP values after 2015 are most likely unaffected (Dizaji and van Bergeijk, 2013). Third, although sanctions' costs are difficult to identify and disentangle from countries' performance, the analysis of export data suggests

---

<sup>10</sup>In fact, Russia is EU's third largest trade partner, and the EU is Russia's largest one.

that only few of our control countries experienced changes in exports that may be attributable to these sanctions. Specifically, a study by Giumelli (2017) finds that, on the one hand, exports of Germany, Italy, France, the Netherlands and Poland seem to be especially hit by the sanctions, with the most severe drop in 2015 compared to 2013. On the other hand, exports of Slovenia, Luxembourg and Romania were affected the least. This brings us to the conclusion that the 2014 sanctions may have had only mild, if any, consequences on the GDP of countries in the control group. Moreover, due to the ambiguity about the direction of the effect in each country, we cannot provide an upper or lower bound for the SC estimator as, e.g. indicated in a similar case by Costalli et al. (2017). Therefore, future research is needed to shed light on the direction and magnitude of sanctions' potential spillovers.

Another source of spillover effects might be changes in military spending in control countries. A strand of literature shows that neighboring states may perceive conflicts as a threat and, thus, increase military expenditures (Smith, 2014; Collier, 2007) which in turn may impact their economic growth (Zielinski et al., 2017; Murdoch and Sandler, 2004). Although similar studies to this commonly assume that indirect spillovers are of negligible magnitude (Costalli et al., 2017; Horiuchi and Mayerson, 2015), we cannot ignore that the Donbass war increased the political instability especially in post-Soviet republics and the Baltic states (Erőss et al., 2016; DeGhett, 2015).

However, also in this case, we believe that changes in military expenditures have only a minor, if any, effect on the results for the following reasons. First, the war's outbreak coincided with multilateral agreements made at the 2014 NATO Summit in which member states were urged to increase their military burden up to 2% GDP share. As a consequence, countries lagging behind this goal made the most significant investments in this sector with both Latvia and Lithuania increasing their military burdens from 0.9% in 2013 up to 1.7% in 2017. The same holds true for Romania and Poland which increased their military burdens over the same period from 1.3% up to 2%, and from 1.8% up to 2%, respectively. Thus, the NATO agreement largely explains the variability of military spending over the years 2014-2017 in Eastern Europe. Secondly, in the case of post-Soviet republics, increased military burden occurs according to bilateral military agreements with the Russian Federation, external military investments from Moscow, and modernization programs within the Collective Security Treaty

Organization (CSTO) which is also led by Russia (Klein, 2019). Third, although military spending may have indirect effects on GDP growth, the size and the direction of this effect is ambiguous. On the one hand, the literature suggests that increased military expenditure has positive effects on economic growth through increased manufacturing output, inciting technological progress and innovation (Barro and Lee, 1994). On the other hand, it can be argued that its effect on investment, capital formation and resource allocation is adverse, indirectly curbing other sectors of the economy and inhibiting long-term economic growth (Knight et al., 1996). Despite the many studies on the topic, there is lack of consensus on the impact of military burden on countries' GDP growth (Herrera and Gentilucci, 2013). As a result, observing no sudden change in the 2014 GDP growth among the control countries, we conclude that the effect of changes in military spending on the donor pool's GDP, if present, is likely negligible. Yet, future research is needed to investigate its magnitude, significance, and spatial dispersion.

## 5 Conclusions

The Donbass war has taken a severe toll on Ukraine, claiming over ten thousand casualties and triggering a severe economic recession. Yet, to the best of our knowledge, there is no empirical evidence on the overall costs incurred by Ukraine as a result of the war. Thus, the goal of this paper is to start filling this gap by estimating Ukraine's GDP foregone due to the Donbass war, and discussing mechanisms underlying this causal effect.

Results from the counterfactual estimation by synthetic controls indicate that the Donbass war led to a considerable decline of Ukraine's economy. Namely, we estimate that, due to this war, the country's per capita GDP decreased by 15.1% (\$1,438.90) on average over the period 2013-2017. Statistical significance of the causal estimates is shown by multiple placebo tests, and robustness is checked by leave-one-out estimations, and confoundedness analyses. In particular, we find that the 2009 gas disputes with Russia and the financial crisis in the same year may lead to overestimated causal effects. As a consequence, the estimated lower-bound of Ukraine's per capita GDP foregone due to the war amounts to 13.89%.

Additionally, we show that the conflict affected the Donbass more severely than the other Ukrainian regions. Over the period 2013-2016, the per capita GRP of the Donbass regions of Donetsk and Luhansk is found to be, on aver-

age, 47% (\$3,825) lower compared to its synthetic counterpart not affected by the strife. This result is in line with the estimated causal effects at country-level.

Several interesting issues are still outstanding. First, this paper focuses on quantifying the economic consequences of the conflict on per capita GDP. Although these account for a large part of direct and indirect costs of the war, we do ignore human capital, social, and psychological effects as well as migration dynamics which start to occur in the longer run. Moreover, given the ongoing nature of the conflict at the time of writing, the continuation of this study should be pursued as more data become available. It should be assessed how the costs evolve over time, in particular, whether the estimated destructive effects increase in scope as more workforce and investment flee the state. This knowledge is crucial to mitigate the damaging consequences of the conflict, and target aid and investment more effectively. These issues are under investigation by the authors.

## Acknowledgements

The paper greatly benefited from discussions at the 2018 conferences of Counterfactual Methods for Policy Impact Evaluation (COMPIE, Berlin), the Annual Congress of the European Economic Association (EEA, Cologne), as well as at the seminars of the Humboldt University, and DIW Berlin. We especially thank J. Paul Elhorst, Markus Eberhardt, Bernd Fitzenberger, and Lorenzo Trapani for fruitful input.

**Declaration of interest statement.** This is to acknowledge that there is not financial interest or benefit that has arisen from the direct applications of our research.

## References

- Abadie, A. (2019). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Article prepared for the Journal of Economic Literature, Technical report, MIT*.
- Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association* 105(490), 493–505.

- Abadie, A., A. Diamond, and J. Hainmueller (2011). Synth: An R package for synthetic control methods in comparative case studies. *Journal of Statistical Software* 42(13).
- Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the Basque country. *The American Economic Review* 93(1), 113–132.
- Ahn, S., Y. Lee, and P. Schmidt (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174(1), 1–14.
- Albalade, D. and G. Bel (2020). Do government formation deadlocks really damage economic growth? Evidence from history’s longest period of government formation impasse. *Governance* 33, 155–171.
- Angelovski, D. (2015). Socio-economic impact and needs assessment. Donbass, Ukraine. *FAO Report*.
- Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2019). Synthetic Difference In Differences. Working Paper arXiv:1812.09970.
- Arunatilake, N., S. Jayasuriya, and S. Kelegama (2001). The economic cost of the war in Sri Lanka. *World Development* 29(9), 1483–1500.
- Åslund, A. (2019). What is wrong with the Ukrainian economy? *Atlantic Council*.
- Athey, S. and G. Imbens (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives* 31(2), 3–32.
- Barro, R. and J.-W. Lee (1994). Sources of economic growth. *Carnegie Rochester Conference Series on Public Policy* 40(1), 1–46.
- Bilmes, L. and J. Stiglitz (2006). The economic costs of the iraq war: An appraisal three years after the beginning of the conflict. *NEBR Working Paper Series* 12054.
- Cao, J. and C. Dowd (2019). Estimation and inference for synthetic control methods with spillover effects. *Working Paper arXiv:1902.07343*.
- Collier, P. (2007). *The Bottom Billion: Why the poorest countries are failing and what can be done about it*. New York: Oxford University Press, Inc.
- Costalli, S., L. Moretti, and C. Pischedda (2017). The economic costs of civil war: Synthetic counterfactual evidence and the effects of ethnic fractionalization. *Journal of Peace Research* 54(1), 80–98.
- de Groot, O., T. Brueck, and C. Bozzoli (2009). How many bucks in a bang: On the estimation of the economic costs of conflict. *Discussion Papers of DIW Berlin* 948.
- DeGhett, T. (2015). Defense and security: Romania is starting to freak out about Russian designs on Transnistria. *Vice News*.
- Deshpande, V. (2018). *Hybrid Warfare: The Changing Character of Conflict*. Pentagon Press.

- Dizaji, S. and P. van Bergeijk (2013). Potential early phase success and ultimate failure of economic sanctions: A VAR approach with an application to Iran. *Journal of Peace Research* 50(6), 721–736.
- Drezner, D. (1999). *The Sanctions Paradox*. New York: Oxford University Press, Inc.
- Echevarría, C. and J. García-Enríquez (2019a). The economic consequences of the Libyan Spring: A synthetic control analysis. *Defence and Peace Economics* 30(5), 592–608.
- Echevarría, C. and J. García-Enríquez (2019b). The economic cost of the Arab Spring: The case of the Egyptian revolution. *Empirical Economics*, 1–25.
- Enders, W., T. Sandler, and G. Parise (1992). An econometric analysis of the impact of terrorism on tourism. *Kyklos* 45(4), 1467–6435.
- Eröss, A., K. Kóvaly, and P. Tátrai (2016). Effects of the Ukrainian crisis in Transcarpathia: The Hungarian perspective. *CMR Working Papers - Center of Migration Research*.
- Ferman, B. and C. Pinto (2017). Placebo tests for synthetic controls. MPRA Paper No. 78079, University Library of Munich, Germany.
- Firpo, S. and V. Possebom (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference, De Gruyter* 6(2), 1–26.
- Foreign Affairs Ministry (2015). Kremlin’s Black Book. Russian war against Ukraine. *Ministry of Foreign Affairs of Ukraine (gov.ua: Ukrainian Cabinet of Ministers)*.
- Gardeazabal, J. (2012). Methods for measuring aggregate costs of conflict. *The Oxford Handbook of the Economics of Peace and Conflict*, 227–251.
- Gates, S. (2012). Development consequences of armed conflict. *World Development* 40(9), 1713–1722.
- Giegerich, B. (2016). Hybrid warfare and the changing character of conflict. *Connections* 15(2), 65–72.
- Giumelli, F. (2017). The redistributive impact of restrictive measures on eu members: Winners and losers from imposing sanctions on russia. *Journal of Common Market Studies* 55(5), 1062–1080.
- Gobillon, L. and T. Magnac (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *The Review of Economics and Statistics* 98(3), 535–551.
- Havlik, P. (2014). Economic Consequences of the Ukraine Conflict. *The Vienna Institute for International Economic Studies, Policy Notes and Reports* 14.
- Herrera, R. and E. Gentilucci (2013). Military spending, technical progress, and economic growth: A critical overview on mainstream defense economics. *Journal of Innovation Economics and Management* 12(2), 13–35.



- Horiuchi, Y. and A. Mayerson (2015). The opportunity cost of conflict: Statistically comparing Israel and synthetic Israel. *Political Science Research and Methods* 3(3), 609–618.
- Hornish, K. (2019). War in the Donbass: Five years on, is there any hope of a durable peace? *The Vienna Institute for International Economic Studies WIIW News*.
- Hussain, Z. (2014). Can political stability hurt economic growth? *World Bank Blog*.
- IEP (2018). Global peace index 2018. *Institute for Economics and Peace (IEP)*.
- Kaempfer, W. and A. Lowenber (2007). The political economy of economic sanctions. *Handbook of Defense Economics* 2, 867–911.
- Keynes, J. (1919). The economic consequences of the peace. *London: Macmillan*.
- Klein, M. (2019). Russia’s military policy in the post-soviet space. *SWP Research Paper*.
- Knight, M., N. Loayzan, and D. Villanuev (1996). The peace dividend: Military expenditure cuts and economic growth. *World Bank Policy Research Working Paper*.
- Kochnev, A. (2019a). Dying light: War and trade of the separatist-controlled areas of Ukraine. *Wiiw Working Paper No. 161*.
- Kochnev, A. (2019b). Smoke on the market: Inattentive investors and the war in Ukraine. *Unpublished manuscript, doi: 10.13140/RG.2.2.20461.33764*.
- Kostanyan, H. and A. Remizov (2017). The donbas blockade: Another blow to the minsk peace process. *CEPS Working Document No 2017/08*.
- Koubi, V. (2005). War and economic performance. *Journal of Peace Research* 42(1), 67–82.
- Kuznetsov, S. (2014). To survive, donbas industry must stay in Ukraine. *Financial Times*.
- Lanoszka, A. (2016). Russian hybrid warfare and extended deterrence in eastern Europe. *International Affairs* 92(1), 175–195.
- Marshall, M. G. (2017). Polity IV project: Political regime characteristics and transitions, 1800-2017. Retrieved from <http://www.systemicpeace.org/inscrdata.html>.
- Murdoch, J. C. and T. Sandler (2004). Civil wars and economic growth: Spatial dispersion. *American Journal of Political Science* 48(1), 138–151.
- OHCHR (2017). Report on the human rights situation in ukraine 16 february to 15 may 2017. *Office of the United Nations High Commissioner for Human Rights*.
- Rácz, A. (2015). Russia’s hybrid war in Ukraine breaking the enemy’s ability to resist. *The Finnish Institute of International Affairs*.

- Reuters (2014). EBRD warns post-Cold War peace dividend at risk. Retrieved from <https://www.reuters.com>.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 5(5), 688–701.
- Skaperdas, S., R. Soares, A. Willman, and S. Miller (2009). The costs of violence. *Washington, DC: World Bank*.
- Smith, R. (2014). The economic costs of military conflict. *Journal of Peace Research* 51(2), 245–256.
- Ukrstat (2014). Ukraine in figures 2014. *Ukrstat - State Statistics Service of Ukraine*.
- Walker, S. (2015). Putin admits Russian military presence in Ukraine for the first time. *The Guardian*.
- Wilson, A. (2015). Ukraine’s economic pressures. *European Council on Foreign Relations*.
- World Bank (2017). Country partnership framework for ukraine. *World Bank Report No. 114516*.
- Zielinski, R., B. Fordham, and K. Schilde (2017). What goes up, must come down? The asymmetric effects of economic growth and international threat on military spendings. *Journal of Peace Research* 54(6), 791–805.

## 6 Appendix

### 6.1 Maps



Figure 9. Ukraine's geographical location in Europe. The darker area in Ukraine represents the Donbass.



Figure 10. Ukraine's administrative division map. Darker areas represent the two Donbass' regions of Donetsk and Luhansk.

## 6.2 Graphs

Figure 11 shows the Leave-One-Out (LOO) estimations: The solid black line represents Ukraine's observed GDP per capita, dashed black line is the original synthetic Ukraine, and gray lines are the LOO synthetic controls estimated by excluding one control unit at the time among those units with positive weights in the synthetic Ukraine (Slovenia, Bulgaria, Armenia, Moldova).

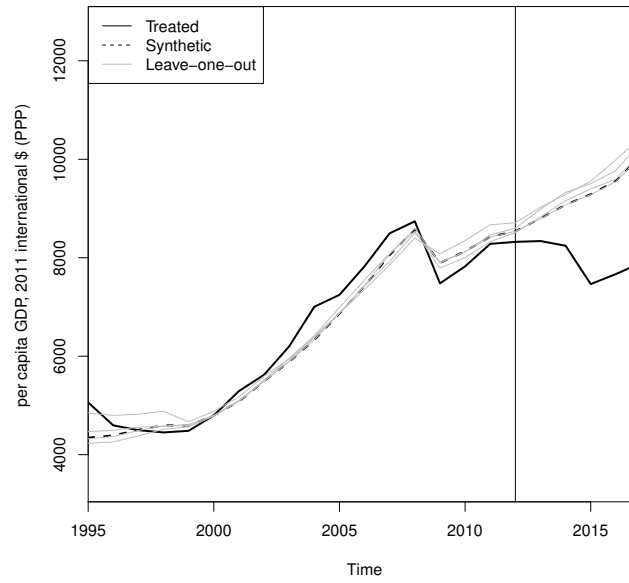


Figure 11. Leave-one-out distribution of synthetic controls for Ukraine.

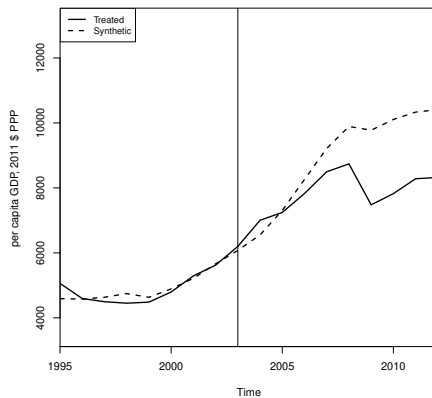


Figure 12. The 2004 Orange Revolution: per capita GDP trends in Ukraine vs. synthetic Ukraine.

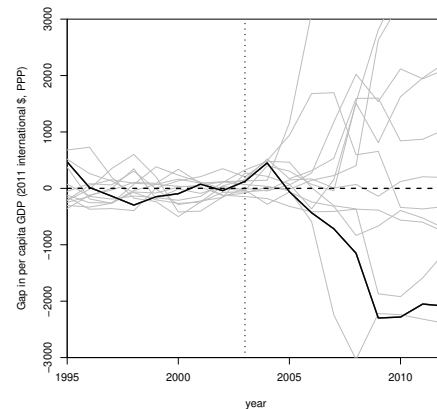


Figure 13. The 2004 Orange Revolution: Gaps in per capita GDP in Ukraine and placebo gaps.

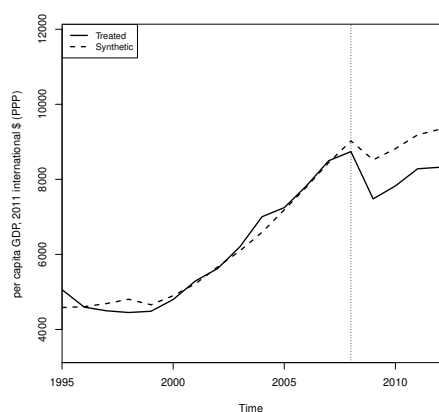


Figure 14. The 2009 gas disputes: Per capita GDP in Ukraine vs. synthetic Ukraine.

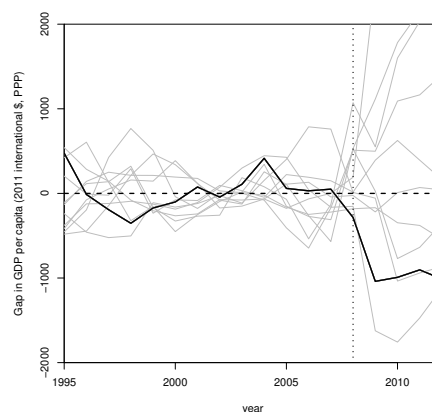


Figure 15. The 2009 gas disputes: Gaps in per capita GDP in Ukraine and placebo gaps.

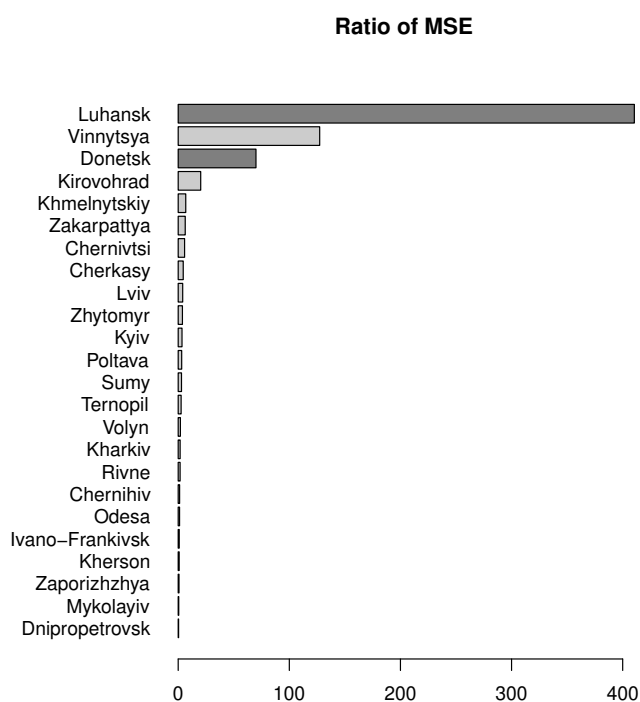


Figure 16. Ratio of mean squared prediction error in post- and pre-war periods for the Donbass regions of Luhansk and Donetsk.

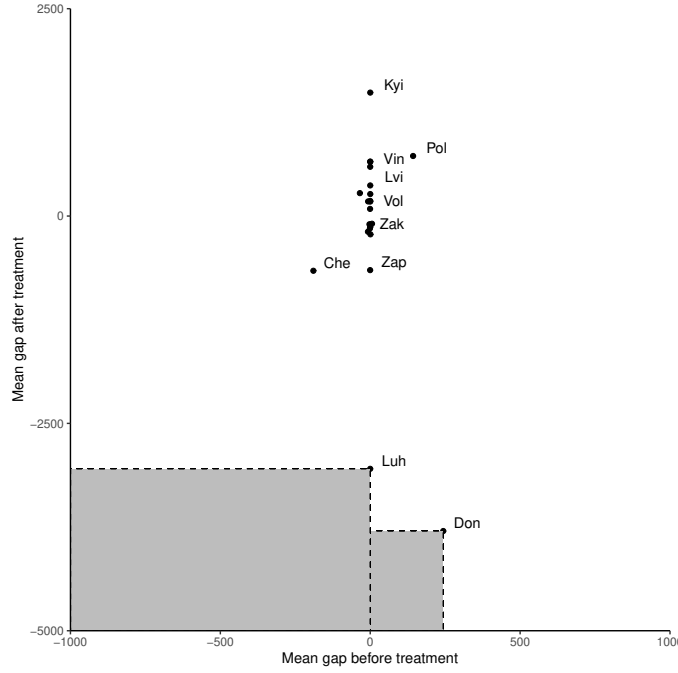


Figure 17. Mean prediction error in post- and pre-war periods (ATT vs. MPE) for the Ukrainian regions. No control unit lies in the highlighted area showing a more extreme ATT as well as a smaller MPE than Luhansk and Donetsk (in absolute terms).

### 6.3 Tables

Table C.1. Data description of variables used in the country-level estimation.

| Variable  | Description   |
|-----------|---|
| Inflation | Inflation, consumer prices (annual %)                             |
| GFCF      | Gross fixed capital formation (% of GDP)                          |
| Polity    | Polity IV Individual Country Regime Trend                         |
| HDI       | Human Development Index   |
| TradeDep  | Sum of exports and imports with the Russian Federation (% of GDP) |

Data source: World Bank's World Development Indicators; Marshall (2017).

Table C.2. Data description of variables used in the regional-level estimation.

| Variable     | Description  |
|--------------|--|
| Export       | Exports of commodities (% of GRP)  |
| Investment   | Capital investment by region (% of GRP)  |
| Unemployment | Unemployment rate of population (results of a sampling survey population of economic activity) |
| GRP          | Per capita gross regional product<br>(2011 international dollars, PPP)                         |

Data source: State Statistics Service of Ukraine (<http://www.ukrstat.gov.ua>).

Table C.3. Donetsk's and Luhansk's control sample with corresponding weights.

| Region          | Donetsk | Luhansk | Region      | Donetsk | Luhansk |
|-----------------|---------|---------|-------------|---------|---------|
| Dnipropetrovsk  | 0.351   | 0.135   | Poltava     | 0.513   | 0.016   |
| Chernivtsi      | 0       | 0.005   | Zakarpattya | 0       | 0.411   |
| Zaporizhzhya    | 0.135   | 0.290   | Odesa       | 0       | 0.006   |
| Volyn           | 0       | 0.005   | Zhytomyr    | 0       | 0.004   |
| Ivano-Frankivsk | 0       | 0.008   | Kyiv        | 0       | 0.021   |
| Kirovohrad      | 0       | 0.005   | Lviv        | 0       | 0.006   |
| Mykolayiv       | 0       | 0.062   | Rivne       | 0       | 0.004   |
| Sumy            | 0       | 0       | Ternopil    | 0       | 0       |
| Kharkiv         | 0       | 0.005   | Kherson     | 0       | 0.003   |
| Khmelnyskiy     | 0       | 0.004   | Cherkasy    | 0       | 0.006   |
| Chernihiv       | 0       | 0       |             |         |         |

Table C.4. Donetsk's per capita GRP predictor means, and Mean Absolute Prediction Error (MAPE) relative to the mean outcome value pre-war.

| Covariate    | Treated | Synthetic | Control sample |
|--------------|---------|-----------|----------------|
| Export       | 25.36   | 18.74     | 9.61           |
| Unemployment | 7.94    | 7.85      | 8.50           |
| Investment   | 18.32   | 21.78     | 23.60          |
| GRP          | 10,091  | 9,847     | 6,014          |
| MAPE         |         | 0.06      |                |

*Note:* All variables are averaged over the 2004-2012 period.

Table C.5. Luhansk's per capita GRP predictor means, and Mean Absolute Prediction Error (MAPE) relative to the mean outcome value pre-war.

| Covariate    | Treated  | Synthetic | Control sample |
|--------------|----------|-----------|----------------|
| Export       | 22.77    | 19.29     | 9.61           |
| Unemployment | 6.90     | 7.89      | 8.50           |
| Investment   | 20.79    | 20.79     | 23.59          |
| GRP          | 6,773.60 | 6,773.70  | 6,013.95       |
| MAPE         |          | 0.02      |                |

*Note:* All variables are averaged over the 2004-2012 period.

Table C.6. Predictor weights in the regional estimations.

| Covariate    | Donetsk | Luhansk |
|--------------|---------|---------|
| Export       | 0.075   | 0.001   |
| Unemployment | 0.733   | 0       |
| Investment   | 0       | 0.408   |
| GRP          | 0.191   | 0.591   |

Table C.7. Exchange rates between UAH and international 2011 dollars.

| Year | Rate  | Year | Rate  |
|------|-------|------|-------|
| 2004 | 0.963 | 2010 | 0.332 |
| 2005 | 0.773 | 2011 | 0.291 |
| 2006 | 0.673 | 2012 | 0.26  |
| 2007 | 0.548 | 2013 | 0.249 |
| 2008 | 0.426 | 2014 | 0.223 |
| 2009 | 0.377 | 2015 | 0.161 |
|      |       | 2016 | 0.132 |

*Note:* Own elaboration based on Ukrainian GDP data obtained from Ukrainian State Statistics Service in UAH and corresponding data obtained from World Bank in international 2011 dollars.



# Erklärung

Zur Erstellung dieser Dissertation habe ich folgende Hilfsmittel genutzt:  
Google, Latex, R, MS Word, MS Excel.

Ich bezeuge durch meine Unterschrift, dass meine Angaben ueber die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, ueber die mir zuteil gewordene Hilfe sowie ueber fruehere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Berlin, den 22. September 2020

Marica Valente